Volume 45 *August 1959* *Number 8*

Proceedings of the
National Academy of Sciences
of the United States of America

The Proceedings of the National Academy of Sciences

OFFICERS OF THE ACADEMY

DETLEV W. BRONK
President

FARRINGTON DANIELS
Vice President

H. P. ROBERTSON
Foreign Secretary

HUGH L. DRYDEN
Home Secretary

WILLIAM J. ROBBINS
Treasurer

EDITORIAL BOARD OF THE PROCEEDINGS

WENDELL M. STANLEY
Chairman

HUGH L. DRYDEN, *Home Secretary*
WILLIAM J. ROBBINS, *Treasurer*

H. P. ROBERTSON, *Foreign Secretary*
EDWIN B. WILSON, *Managing Editor*

GREGORY BREIT	THOMAS FRANCIS, JR.	H. W. MAGOUN
HARRISON BROWN	C. H. GRAHAM	HARLOW SHAPLEY
ERNST CLOOS	SAUNDERS MAC LANE	PAUL WEISS

Subscription price is $12.50 for 1 year, $24.00 for 2 years, and $34.50 for 3 years; the price of a single issue is $1.50. Make all remittances payable to the National Academy of Sciences in U.S. currency or its equivalent.

Editorial correspondence should be addressed to the PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, 2101 Constitution Avenue, Washington 25, D. C.

Business correspondence should be addressed to the Printing and Publishing Office, National Academy of Sciences, 2101 Constitution Ave., Washington 25, D. C.

Subscribers are requested to notify the Printing and Publishing Office of the Academy and their local postmaster immediately of change of address. Notices to the Academy should provide both the old and the new address.

Microfilms of complete volumes of this journal are available to regular subscribers only and may be obtained at the end of the volume year from University Microfilms, 313 N. First Street, Ann Arbor, Michigan.

Second-class postage paid at Easton, Pennsylvania. This publication is authorized to be mailed at the special rates of postage prescribed by Section 132.122.

PRINTED IN THE U. S. A.

THE PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES
is published monthly by
THE NATIONAL ACADEMY OF SCIENCES

Proceedings of the

NATIONAL ACADEMY OF SCIENCES

Volume 45 · Number 8 · August 15, 1959

SYMPOSIUM ON SCIENTIFIC EFFECTS OF ARTIFICIALLY INTRODUCED RADIATIONS AT HIGH ALTITUDES

BY INVITATION OF THE COMMITTEE ON ARRANGEMENTS FOR THE ANNUAL MEETING

Read before the Academy, April 29, 1959

CONTENTS

CHAIRMAN'S INTRODUCTORY REMARKS . . . . . *Richard W. Porter* 1141
THE ARGUS EXPERIMENT . . . . . *N. C. Christofilos* 1144
SATELLITE OBSERVATIONS OF ELECTRONS ARTIFICIALLY INJECTED INTO THE GEOMAGNETIC FIELD . . . *James A. Van Allen, Carl E. McIlwain, and George H. Ludwig* 1152
PROJECT JASON MEASUREMENT OF TRAPPED ELECTRONS FROM A NUCLEAR DEVICE BY SOUNDING ROCKETS . . *Lew Allen, Jr., James L. Beavers, II, William A. Whitaker, Jasper A. Welch, Jr.. and Roddy B. Walton* . 1171
THEORY OF GEOMAGNETICALLY TRAPPED ELECTRONS FROM AN ARTIFICIAL SOURCE . . . . . *Jasper A. Welch, Jr., and William A. Whitaker* 1190
OPTICAL, ELECTROMAGNETIC, AND SATELLITE OBSERVATIONS OF HIGH-ALTITUDE NUCLEAR DETONATIONS—PART I . . . . . *Philip Newman* 1208
OPTICAL, ELECTROMAGNETIC, AND SATELLITE OBSERVATIONS OF HIGH-ALTITUDE NUCLEAR DETONATIONS—PART II . . . *Allen M. Peterson* 1221

CHAIRMAN'S INTRODUCTORY REMARKS

BY RICHARD W. PORTER

CHAIRMAN, TECHNICAL PANEL ON THE EARTH SATELLITE PROGRAM, U. S. NATIONAL COMMITTEE FOR THE INTERNATIONAL GEOPHYSICAL YEAR 1957–1958

The existence of corpuscular radiation of high intensity in the vicinity of the earth was first discovered by accident in the course of the first successful experiment of the United States IGY satellite program. The scientific equipment on our first satellite, 1958 Alpha, or—as it was popularly called—Explorer I, had been designed to measure the intensity of primary cosmic rays entering the upper layers of the earth's atmosphere. Although this equipment was capable of detecting and counting a number of ionizing events well in excess of any expected cosmic ray intensity, the telemetered recordings showed that the satellite was passing in and out of regions where the number of counts rose abruptly to a level where it not only saturated the

counting circuits, but even the Geiger-Mueller tube itself. Fortunately, by means of laboratory tests on similar instruments, it was possible to identify, at least approximately, the levels at which this saturation occurred and thereby to establish a lower limit for the magnitude of the radiation that was being encountered.

The data from this first US-IGY satellite, launched on January 31, 1958, by the Army Ballistic Missile Agency and the Jet Propulsion Laboratory of the California Institute of Technology, and also from the next IGY Explorer, 1958 Gamma, launched on March 26, 1958, again by ABMA and JPL, led to the following preliminary conclusions which were presented by Dr. Van Allen at an historic joint meeting of members of the National Academy of Sciences and the American Physical Society here in Washington a year ago:

(1) Below approximately 700 km, at latitudes between 33 degrees north and south, the satellites were encountering only radiation that was essentially in accord with expectations on the basis of previous cosmic ray measurements in rockets and balloons and at the earth's surface.

(2) At higher altitudes, generally around 1000 km or above, depending somewhat on longitude and latitude, the radiation intensity began to increase very rapidly with altitude in a manner that could not be explained on the basis of cosmic ray phenomena.

(3) At the highest altitudes, on the assumption that the radiation consisted entirely of penetrating charged particles, that is, of electrons having an energy of at least 3 Mev or protons having at least 30 Mev, the omnidirectional intensity was found to exceed 1700 counts per square centimeter per second. On the other hand, if the energy were assumed to consist exclusively of electrons in the range between approximately 50 Kev and 3 Mev, where the electrons themselves could not penetrate the absorber but could produce Bremsstrahlung energetic enough to penetrate without much attenuation, the omnidirectional intensity must have been in excess of 10^8 per square centimeter per second.

(4) Because of the rapid change in observed intensity between 700 and 1,000 km, where the air density is insignificantly small, it was evident that the primary radiation must be restrained from reaching lower altitudes by the earth's magnetic field and therefore must consist of charged particles having energies substantially less than those normally associated with cosmic rays.

On the basis of these conclusions, it was suggested that the observed radiation might originally have come from the sun in the form of ionized gas clouds, portions of which subsequently entered the earth's magnetic field by means of local magnetic disturbances created by the clouds themselves. Having once entered the earth's field, the particles might remain trapped in a manner suggested previously by Stoermer and others. It was, therefore, quite natural to relate this radiation to the "soft auroral radiation" which had been observed in the arctic and antarctic regions, both before and during the IGY. Both Van Allen and Bennett had suggested that observations be made of auroral corpuscular radiation using satellites having highly inclined orbits. It had not been anticipated, however, that this radiation would be found at such low altitudes so near the equator. Nevertheless, on the basis of this previous experience there was a tendency to favor the alternative of Bremsstrahlung from low-energy electrons as the predominant radiation source.

The National Academy of Sciences' IGY Committee, and particularly its Satel-

lite Panel, were anxious to conduct further experiments as soon as possible with instrumentation more suitably designed to observe this newly-discovered natural radiation belt. Work was started immediately to prepare instrumentation and the Department of Defense was asked to provide at least two more Jupiter C vehicles.

One of the two additional satellites, Explorer IV, was launched by ABMA and JPL on July 26 and became 1958 Epsilon; the other failed to go into orbit.

While Explorer IV was still in the planning stage, the Advanced Research Projects Agency of the Department of Defense carried forward plans for the Argus experiment which we are here to talk about this morning. Inasmuch as the objective of the Argus experiment was to observe the behavior of electrons trapped in the earth's magnetic field, instrumentation was provided for Explorer IV which was suitable for study of both the natural radiation belt and any Argus effects. Pertinent scientific results relating to the natural radiation effects observed by this satellite were presented at the American Physical Society meeting in Chicago earlier this year and were published in the March issue of the *Journal of Geophysical Research*.

Despite some newspaper stories to the contrary, there was no commitment under the IGY to exchange information with other nations on the Argus experiment. Argus was not an IGY program; it was a Department of Defense effort. Some confusion on this point has arisen entirely from the circumstance that a single satellite gathered data on two subjects—one, natural radiation phenomena; the other, artificial phenomena from Argus. However, the Department of Defense and the Atomic Energy Commission have now released the information you are about to hear this morning so that it can be made a part of the synoptic geophysical record of this interesting period in science.

Presentation of these results necessarily waited upon reduction and analysis of the scientific data and upon declassification of the Argus experiment.

The first Argus burst produced a yield of between 1 and 2 kilotons at approximately 02 hr 30 min UT on the 27th of August at a nominal altitude of 480 km in the neighborhood of longitude 12° west, latitude 38° south. A fascinating sequence of observations followed which will be described in some detail by the speakers on this morning's program. For the first time in history, geophysical phenomena on a world-wide scale were being measured and related to a quantitatively known cause —namely, the injection into the earth's magnetic field of a known quantity of electrons of known energies at a known position and at a known time.

The diverse radiation instruments in Explorer IV recorded and reported to ground stations the absolute intensity and position of this man-made shell of high energy electrons as the satellite continued to lace back and forth through the trapped radiation, hour after hour and day after day. The rate of decay of electron density as a function of altitude provided new information on the density of the remote upper atmosphere since atmospheric scattering was the dominant mechanisms for loss of particles. Moreover, the continuing observation of the thickness of the shell served to answer the vital question as to the rate of diffusion of trapped particles transverse to the shell. All of these matters, which are of essential importance to a thorough understanding of the electrodynamics of the natural radiation, thus became subject to direct study by means of the electrons released from Argus I.

At approximately 03 hr 20 min UT on August 30, a second detonation of the same size was produced, also at nominal altitude of 480 km in the neighborhood of longi-

tude 8° west, latitude 50° south. A similar succession of phenomena followed.

The third and final detonation, of the same size, was at about 22 hr 10 min UT on September 6 at the same nominal altitude in the neighborhood of longitude 10° west, latitude 50° south. Again a series of remarkable geophysical phenomena was produced. And for the third time, Explorer IV faithfully observed the power of artificially-produced trapped radiation.

Throughout the testing period a series of firings of high altitude sounding rockets was carried out successfully, yielding valuable results in the lower fringes of the trapping region.

Explorer IV continued to observe the artificially-injected electrons from the Argus tests, making some 250 transits of the shell, until exhaustion of its batteries in the latter part of September. By that time the intensity had become barely observable above the background of natural radiation at the altitudes covered by the orbit of this satellite. It also appears likely that the deep space probe Pioneer III detected a small residuum of the Argus effect at very high altitudes on December 6, 1958, but the effect appears to have become unobservable before the flight of Pioneer IV on March 3, 1959.

An immense body of directly related observations has now been under study and interpretation by a large number of persons for about seven months and, as you will shortly hear, useful scientific results are being derived. It is also possible that other important contributions will arise as the many diverse geophysical observations being conducted by other countries participating in the International Geophysical Year are accumulated and analyzed.

---

*THE ARGUS EXPERIMENT**

By N. C. Christofilos

LAWRENCE RADIATION LABORATORY, UNIVERSITY OF CALIFORNIA, LIVERMORE

With the advent of the satellite era the thoughts of the scientists have extended to outer space, since means have become available for measurements and experimentation in the region of the upper atmosphere. Two classes of experiments are possible with satellite-carried instruments:

(1) To measure natural phenomena.
(2) To create artificial effects under controlled conditions.

This class of experiments is most interesting as the initial conditions are known. Hence physical quantities can be measured under completely different conditions than those which the natural phenomena allow.

Just after the first Sputnik was launched I was attracted by outer space problems. The second class of experiments appeared very intriguing to me and I started thinking about possible experiments in outer space. The first experiment which occurred to me stemmed out of my main field of endeavor in the last five years, namely, the Astron thermonuclear device. In this device it is contemplated to establish a cylindrical layer of relativistic electrons between two mirror coils.

This electron layer then will establish in turn a magnetic bottle for plasma confinement, while the relativistic electrons will subsequently heat this plasma by Coulomb collisions up to fusion temperature. Then it occurred to me to extrapolate this idea of the electron layer to global dimensions, inasmuch as the earth's magnetic field provides a natural mirror effect. The trapping phenomenon was known theoretically for many years from the work of Stoermer, Alfven, and others, namely, that charged particles released at one point will spiral about a magnetic line being reflected at two points along this line. These points came to be known as mirror points. In addition to this motion the electrons would drift eastward due to the fact that the earth's field is decreasing as one moves away from the earth. Consequently, it was obvious that if sufficient electrons would be released at one point, these electrons would spread all around the earth creating a layer of rather small thickness but of global dimensions. Thus the first conclusion was that the electrons could be released from one point source; provided, however, that this source would be sufficiently large.

The question then arose whether or not these electrons will survive long enough to complete several circuits around the earth. The time for one circuit around the earth is approximately inversely proportional to the electron energy. The lifetime has been calculated. The conclusions from these calculations was that most of the losses are due to Coulomb scattering with air atoms. Since the density of the atmosphere decreases exponentially with altitude it turned out that most of the losses occur near the mirror points. The result of the scattering is that the mirror points move down toward denser atmosphere. The electrons are practically lost after the mirror points have moved downward by about one scale height. The lifetime is proportional to the air density at the mirror points and approximately to the square of the electron energy, and is proportional to the length of the magnetic line. The conclusion was that electrons of one to two mev energy injected at an altitude of a few hundred miles will survive several hours or more, depending on the air density at the mirror points. In more detail the results of these calculations were (see also Appendix 1):

The energy loss ($\dot{E}$) of the electrons by Coulomb scattering is equal to the energy loss ($\dot{E}_0$) corresponding to the air density at the mirror points divided by the length (ψ) of the magnetic line from the mirror point to the equatorial plane, measured in units of earth radii, times the square root of the ratio of the earth's radius (r_0) to the scale height (h). Namely

$$\dot{E} = \dot{E}_0/(\psi\sqrt{r_0/h}) \tag{1}$$

or

$$\dot{\gamma} = 3\cdot 10^{-8}\, N_0 Z/(\psi\sqrt{r_0/h}) \qquad \text{rest mass units/day} \tag{2}$$

From the energy loss we can compute the square of the scattering angle and thus the downward velocity of the mirror points.

$$\dot{r}/r_0 = -\frac{Z(Z+1)}{\gamma^2}\,\frac{10^{-8}\,N_0}{\psi}\sqrt{\frac{h}{r_0}} \qquad \text{days}^{-1} \tag{3}$$

where Z is the atomic number, N_0 the numerical density, r_0 the earth radius, and γ the electron energy in rest mass units.

Electrons in the few mev range are lost by scattering without appreciable energy loss. In this case the lifetime is

$$T = 1.7 \cdot 10^6 \frac{\psi \gamma^2}{N_0} \sqrt{\frac{h}{r_0}} \quad \text{days} \tag{4}$$

For the following parameters

$$(r_0/h)^{1/2} = 8 \qquad N_0 = 10^4$$
$$\psi = 1.5 \text{ (latitude } 45°) \qquad \gamma = 3$$
$$T = 2.86 \text{ days}$$

These results pointed toward a convenient source of a large quantity of electrons, namely, to an A-bomb. This source is so plentiful that even one megaton of fission would create an electron layer so dense as to constitute a radiation hazard in outer space. To illustrate this I cite the following example: one megaton of fission yields 10^{26} fissions, approximately. If we assume that 4 electrons per fission are above 1 mev energy and half will be trapped in the earth's field we derive the number of trapped electrons, namely, $2 \cdot 10^{26}$. Then let us assume that these will be spread in a volume in outer space equal to the earth's volume, or 10^{27} cm.^{3} The resulting electron density is 0.2 electron/cm.^{3} The flux against any surface exposed to the electrons is

$$N_e = n_e c/4 \tag{5}$$

or

$$N_e = 1.5 \cdot 10^9 \quad \text{electrons/cm}^2\text{, sec.}$$

A small part of the energy of the electrons is converted to bremsstrahlung. The efficiency of this conversion is proportional to the atomic number of the target material.

The radiation level inside a space vehicle is

$$R = 8Z\gamma(\gamma - 1)\, n_e \quad \text{roentgens/hour} \tag{6}$$

where Z and γ were defined previously. Assuming an atomic number $Z = 10$ and $\gamma = 3$, we observe that the radiation level is more than 100 roentgens/hour, which is a good fraction of the lethal dose. Consequently, it is obvious that any explosion of such magnitude can create a radiation hazard in outer space and any space experiments involving A-bomb explosions must be carefully designed to avoid creation of hazardous radiation.

Fortunately a much smaller yield, namely in the kiloton range, is sufficient to yield detectable quantities without creating any radiation hazard at all.

It is obvious that the scientific aspect of such experiments is extremely important. Consequently, as soon as I suggested this experiment about the end of October, 1957, the matter was discussed with interest within the Radiation Laboratory for several weeks. In January, 1958, I had completed preliminary calculations and a paper was prepared. The matter was further discussed with scientists from other laboratories. Finally, a group of scientists was appointed to evaluate the proposal, and it was recommended to proceed with a test employing a small yield A-bomb in the kiloton range. Further an altitude was selected such so that the effect would

not last more than a few days. In the meantime the natural radiation belt was discovered by Dr. James Van Allen.

This discovery was partial proof that the trapping effect actually exists as postulated, so that about the end of April, 1958 it was decided to proceed with the Argus test. Argus was the code name given to the experiment. The most pertinent question then was the selection of the location of the tests. Since the magnetic axis of the earth is tilted and the center of the dipole is displaced toward the Pacific Ocean, it is obvious that the trace of the mirror points will have their closest distance from the surface of the earth over the South Atlantic Ocean. Hence a location in the South Atlantic was finally selected for the test.

The over-all responsibility for the preparation and conduct of the experiment was vested in the Advanced Research Projects Agency (A.R.P.A.), then just established. It required only four months from the time it was decided to proceed with the tests until the first bomb was exploded. This is a remarkable achievement in organization, and if we were to single out the one man as most responsible for this accomplishment it would be Dr. Herbert York, who was at that time Chief Scientist of the A.R.P.A. During this short time a task force including the rocket launching ship, Norton Sound, was organized, and special instrumentation on satellites was prepared by Dr. James Van Allen. Many other activities were coordinated with the help of dozens of scientists.

In addition to the trapping effect the creation of an artificial aurora was predicted. The nontrapped electrons would be guided along the magnetic lines of force (which pass through the burst point) toward the upper atmosphere, producing strong ionization along these magnetic lines. Hence in both points where these lines re-enter the atmosphere auroral luminescence was expected.

The persons responsible for organizing the observations of auroral phenomena and magnetic disturbances as well as some of the minitrack stations were Dr. Philip Newman (A.F.C.R.C.) and Dr. Allen Peterson (S.R.I.). Besides the satellite observations, some rocket launchings were scheduled for observations at the very beginning of the formation of the Argus shell. This responsibility was assumed by a group of scientists in the A.F.S.W.C.

The experiment was successful in all respects except that Explorer V was not placed in orbit. However, the one satellite available, Explorer IV, yielded most of the important information although valuable information resulted from other observations.

The bursts occurred on the 27th and 30th of August in the early morning hours and on the 6th of September shortly before midnight, Greenwich time. The locations were, respectively, 38°S, 12°W; 50°S, 8°W; and 50°S, 10°W, approximately.

A very interesting sequence of observations was obtained. The initial flash of the burst was followed by an auroral luminescence extending upward and downward along the magnetic line where the burst occurred. Simultaneously at the point where the same magnetic line returns to the earth's atmosphere (the so-called conjugate point) in the North Atlantic another auroral luminescence appeared in the night sky near the Azores Islands.

The instruments of Explorer IV recorded and reported to ground stations, around the earth, the electron density in the shell, as well as the position and its thickness. The satellite continued to penetrate the shell several times a day at

various altitudes, so that the density as a function of altitude and time was measured. The apogee of the electrons was about an earth's radius out in space or 4000 miles approximately. Thus, for the first time the earth's magnetic field was being plotted experimentally. The conjugate points were a few hundred miles farther away than the theoretical calculations indicated. However, this is rather a small error considering that ring currents at high altitudes are not yet well known and that all the measurements of the earth's field were thus far taken at the earth's surface.

When the location in the South Atlantic was selected, the data on the natural radiation belt were rather meager. However, it turned out that the Argus shell was located at a region where the intensity of the natural radiation passes through a minimum. If the location were further south or north of the one selected, the Argus shell effects would have disappeared in the background much earlier. For the first time an experiment was conducted in outer space on a global scale where all the measured quantities were related to a known cause, namely, to the trapping in the earth's field of a known number of electrons of a known energy, injected at a known location at a known time.

The satellite measurements are described in detail by Dr. James Van Allen.[1] The general conclusion on electron lifetime is that the electrons died of natural death, namely, by scattering with the air molecules near the mirror points. No effect was observed which appeared to expedite or enhance the loss of the electrons.

Another important observation was that (within the limit of observational error) neither did the electron shell move at all across the magnetic lines nor did the electrons diffuse across these lines. This verifies the theory that there are two adiabatic invariants, namely, the magnetic moment and the action or the integral of the momentum along the magnetic lines. The first adiabatic invariant has been known for a long time. The second was postulated in 1953 by Marshall Rosenbluth. According to a theory by T. Northrup and E. Teller,[2] the particles follow a closed magnetic surface even if the field is not axially symmetric. A rigorous proof that these quantities are invariants in all orders has been provided by Dr. Martin Kruskal (Princeton University), the first about two years ago, the other recently. Also it was believed that invariance in time exists. This means that any change of the magnetic field in time in a period which is long in comparison with the period of one oscillation between mirror points will not affect permanently the position of the shell.

Since the changes in the earth's field occur in a matter of hours, very rarely in minutes, whereas the oscillation period is less than a second, one should expect no effect in the position of the shell. Actually these postulations have been verified. In connection with the position stability of the shell another conclusion can be drawn concerning hydromagnetic instabilities. According to hydromagnetic theory the relevant quantity in generating unstable motions is the ratio of the particles' energy density to the magnetic field energy density. In the case of the Argus shell this ratio was negligible; hence the trapped electrons could not create any instability. However, an ionized plasma either permanently existing there or temporarily trapped as a result of the sun's activity would create such motions characterized by unstable magnetic line interchange. In such a case the Argus shell would have followed the motion of the magnetic lines and be deformed and

change location. Since no such motion has been observed this is a good indication that such instabilities are not occurring in the earth's field at least at the latitudes of the Argus shell or lower (toward the equator). This is a useful observation and I will elaborate later on some possible conclusions.

The first Argus experiment provided us with very important information. However, this is only a fraction of what is possible with experiments of this nature. A variety of experiments is possible as well as a variety of methods of injecting electrons into the geomagnetic field. We can classify the means of injection into three categories:

(*a*) Injection through atomic explosions as in the Argus experiment.
(*b*) Injection from a satellite carrying a payload of radioactive β-decay material.
(*c*) Injection from a satellite-borne electron accelerator.

Comparing the three categories, we observe that the first is the easiest, but the last affords much more controlled experiments since monoenergetic electrons can be injected in any desirable direction. However, experiments of the first category are already being conducted whereas a satellite-borne accelerator might require two years to be developed and launched. Consequently, one can plan a sequence of experiments over a number of years in such a way that the satellite-borne accelerator experiments will be scheduled two to three years hence. In such a sequence of experiments several important results of geophysical interest may be obtained. Some of these are:

(1) Complete mapping of the earth's magnetic field and defining the extent of the region of closed lines of the earth's field. According to existing theories this region ends at a distance from the earth, at the equatorial plane, where the magnetic pressure equals the gas pressure. Since this distance is many earth radii, the residual gas there is the solar corona consisting of hydrogen atoms in an ionized state. The equivalent gas pressure is half the mass density times the square of the orbital velocity of the earth. Due to this effect, according to these theories, the earth's magnetic field is swept away beyond this distance. A magnetic line passing through the equatorial plane at a distance of 10 radii returns to the earth at 70° north or south latitude. This is the latitude of the auroral zones. Hence, it is quite possible that lines originating in the polar regions beyond this latitude do not return to the earth but vanish in outer space intermixing with the sun's field or the interplanetary field. It is of extreme interest to determine the earth's field in this fringe region. It is quite probable that this region is very turbulent and that the region of the closed magnetic lines is variable during magnetic storms or other disturbances created by the interaction of plasma streams originating in the sun and the earth's magnetic field. By conducting well designed Argus experiments it is possible not only to monitor these changes but also to measure approximately the density of the solar corona near the earth. Further, it will be possible to clarify auroral phenomena and advance the theory of such phenomena.

(2) To shed light on the origin and lifetime of the natural radiation belts. The first indications are that the region of the outer belt is very turbulent. Hence we cannot exclude the possibility that particles which originated in the sun are responsible for the outer belt. A continuous monitoring of the density of artificially injected electrons in that region might shed light not only on the origin of the outer

radiation belt but on the origin and behavior of the auroral phenomena as well.

The lifetime of the radiation belts cannot be directly determined since the density is more or less constant under steady-state conditions. However, by injecting labeled electrons and measuring the lifetime we can determine the lifetime of the particles in the belt. Consequently, the required source strength will be derived. Thus it will be easier to explain the origin of the natural radiation belts. As a matter of fact, already one observation from the Argus experiment appears to be very useful and offers tentative conclusions on the origin of the inner belt. The Argus shell was located beyond the maximum of the inner belt. As I mentioned above, this shell exhibited remarkable stability and did not move at all during the few weeks when observations were possible. This indicates that no convection currents existed. Consequently, plasma cannot be transferred from the outside to the inner belt. However, since during the Argus experiment there did not occur any extremely violent solar disturbances, this conclusion must be considered as being only tentatively true at this time. Then, according to this tentative conclusion, the inner belt particles are born within this region. The only explanation therefore is that the inner belt particles are neutron decay products where the neutrons originate from cosmic ray interactions with the upper atmosphere. It seems that this radiation consists of protons and electrons, both products of neutron decay. During the initial discussion on Argus, I suggested the possibility of neutron decay and I postulated that electrons must already be trapped up there. However, I did not proceed with the quantitative calculations, which have been done later by Dr. Kellog.

In suggesting this effect I have in mind thermal neutrons. Professor Singer, however, proposed that high energy neutrons are responsible, yielding upon decay high energy protons. These theories of neutron decay origin seem to be verified by the stability of the Argus shell. However, final conclusions will be possible only after further experiments, especially, an Argus experiment during a period of intense solar activity.

The above result was not expected when the Argus experiment was first proposed since the natural radiation belt was not yet discovered. Continuation of the Argus experiments in any of the three categories may yield new information in addition to that anticipated. This has happened many times in physical experiments, particularly when a new class of experiments was conducted in a relatively new field.

In conclusion, I would like to say that there is now available a new tool for exploring the part of outer space encompassed by the geomagnetic field, and for clarifying the phenomena of interaction of the earth's magnetic field with various charged particles of natural or artificial origin.

APPENDIX

Calculation of the Lifetime of the Electrons.—The electron loss is due mainly to Coulomb scattering near the mirror points. A given increase ($\Delta\theta^2$) of the square of the scattering angle will cause an increase of the square of the parallel velocity by $\Delta W_{||}{}^2$, namely

$$\Delta W_{||}{}^2 = c^2 \Delta\theta^2 \tag{1}$$

From the invariance of the magnetic movement we have the relation

$$W_{||}^2 = c^2\left(1 - \frac{B}{B_m}\right) \tag{2}$$

where

$$\frac{B}{B_m} = \left(\frac{r_0}{r}\right)^3 \tag{3}$$

r_0, B_m are the radius and field at the mirror point, respectively. Equation (3) is valid at high latitudes where the lines are almost normal to the earth's surface.

Equations (2) and (3) yield

$$W = \dot{R} = c\left[1 - \left(\frac{r_0}{r}\right)^3\right]^{1/2} \tag{4}$$

From the above equations we derive

$$\Delta R = \frac{r_0}{3}\,\Delta\theta^2 \tag{5}$$

where ΔR the downward shift of the mirror points caused by an increase of the square of the scattering angle by $\Delta\theta^2$.

By differentiating equation (5) we obtain

$$\dot{R} = \frac{r_0}{3}\,\dot{\theta}^2 \tag{6}$$

The value of $\dot{\theta}^2$ is given by the known equation

$$\dot{\theta}^2 = -\frac{(Z+1)\dot{\gamma}}{\gamma^2} \tag{7}$$

where $\dot{\gamma}$ is the energy loss by scattering measured in rest mass units, (γ is the electron energy in rest mass units). The energy loss by scattering is given by the known Bethe formula. In the present case, after substituting the numerical values we obtain

$$\dot{\gamma} = -3.10^{-8}\,NZ \qquad \text{rmu/day} \tag{8}$$

The value of the air density is a function of position which in turn is a function of time. We shall express the air density as a function of time and integrate along a trajectory from the point of reflection to the equational plane.

By integrating equation (4) we obtain

$$\frac{ct}{r_0} = \int \frac{d(r/r_0)}{\left[1 - \left(\frac{r_0}{r}\right)^3\right]^{1/2}} \tag{9}$$

Since most of the losses occur at a distance $(r - r_0)$ from the mirror points much less than r_0 we can substitute, $x = r - r_0$, in equation (9), which upon integration yields

$$x = \frac{3}{4}\frac{c^2}{r_0}\,t^2 \tag{10}$$

The air density varies with the distance (x) from the mirror points as $N_0 e^{-x/h}$ where N_0 is the air density and h the scale height at the mirror points, respecitvely. From the above equations we obtain

$$\dot{\gamma} = -3.10^{-8} N_0 Z \exp\left(-\frac{3c^2}{4r_0 h} t^2\right) \tag{11}$$

Integration of this equation from zero to infinity shall yield the energy loss along one trajectory from the reflection point to the equator. Then considering the oscillation time we derive the average energy loss, namely

$$<\dot{\gamma}> = \frac{3.10^{-8} N_0 Z}{\psi} \sqrt{\frac{h}{r_0}} \quad \text{rmu/day} \tag{12}$$

where $(r_0\psi)$ is the length of the magnetic line from the mirror point to the equator. The rate of change of the scattering angle ($\dot{\theta}^2$) is a function of $\dot{\gamma}$, just evaluated, and a function of γ^2. The value of γ is reduced from Coulomb scattering and Schwinger radiation loss. However, the contribution from Schwinger radiation is small for N_0 more than 10^5 atoms per cm^3. Equations (6), (7), and (12) yield

$$\dot{x} \equiv \dot{R} = -\frac{10^{-8} Z(Z+1) N_0}{\psi\gamma^2} r_0 \sqrt{\frac{h}{r_0}} \tag{13}$$

As the mirror points move downward the air density increases. Consequently at a distance x (downward) from the initial mirror point the downward velocity is

$$\dot{x} = -\frac{10^{-8} Z(Z+1)}{\psi\gamma^2} N_0 e^{-x/h} r_0 \sqrt{\frac{h}{r_0}} \tag{14}$$

In the case of electrons of a few Mev energy, the loss of the electrons is due more to scattering than to energy loss, namely, the mirror points move downward one scale height without appreciable energy loss. In this case equation (14) yields upon integration

$$T = 1.7 \cdot 10^6 \frac{\psi\gamma^2}{N_0} \sqrt{\frac{h}{r_0}} \text{ days.} \tag{15}$$

* Work was performed under auspices of the U. S. Atomic Energy Commission.

[1] Van Allen, James A., Carl E. McIlwain, and George H. Ludwig, these PROCEEDINGS, **45**, 1152–1171 (1959).

[2] Northrop, T., and E. Teller, Univ. of Calif. Rad. Lab. Report, UCRL-5615 (to be published).

SATELLITE OBSERVATIONS OF ELECTRONS ARTIFICIALLY INJECTED INTO THE GEOMAGNETIC FIELD

BY JAMES A. VAN ALLEN, CARL E. MCILWAIN, AND GEORGE H. LUDWIG

DEPARTMENT OF PHYSICS, STATE UNIVERSITY OF IOWA

Following our discovery with Explorer I (Satellite 1958 Alpha) and with Explorer III (Satellite 1958 Gamma) that there were very great intensities of charged particles trapped in the geomagnetic field,[1] we undertook to make arrangements for a further satellite flight of equipment of greater discrimination and much greater

dynamic range for the purpose of detailed study of the properties of the radiation and of its spatial distribution. The progress of such arrangements was greatly aided by Richard W. Porter, chairman of the Technical Panel for the Earth Satellite Program of the U.S. National Committee for the International Geophysical Year, and by Herbert York, then chief scientist of the Advanced Research Projects Agency of the Department of Defense.

In October, 1957, Nicholas C. Christofilos of the Lawrence Radiation Laboratories of the University of California at Livermore had proposed in an unpublished memorandum that many observable geophysical effects could be produced by an atomic detonation at high altitude above the earth in the tenuous upper atmosphere. Of the various effects contemplated one of the most interesting promised to be the temporary trapping of high energy electrons at high altitudes in the geomagnetic field. Such electrons result from the radioactive decay of fission fragments and, less importantly, of neutrons.

A subsequently organized study group carried out more detailed estimates of the various effects to be expected and concluded that the proposed experiments (later named "Argus" experiments) were indeed feasible. This study group, led by W. K. H. Panofsky, also concluded that satellite observations would be of great value.

Meanwhile, the discovery of the existence of the natural, trapped radiation served as an over-all validation of the Argus proposal in that it showed that a high intensity of charged particles can indeed be trapped in the geomagnetic field. And to the I.G.Y. workers at the State University of Iowa and elsewhere it was clear that the artificial, impulsive injection of known particles along a known line of force, at a known time, would be a powerful technique for elucidating many of the uncertain aspects of the dynamics of geomagnetic trapping.

After a series of conferences, the Advanced Research Projects Agency agreed to provide, among other things, two Jupiter C satellite vehicles for the joint I.G.Y.-Argus undertaking.

Design, development, construction, and calibration of the detectors, associated electronics, etc., were accomplished within the Department of Physics of the State University of Iowa (S.U.I.). The apparatus was designed around the two-fold purpose of further, detailed study of the natural radiation and of detection and quantitative study of artificially injected electrons from the Argus experiments. Final environmental tests and over-all specifications of the flight payloads were in the hands of the Army Ballistic Missile Agency (ABMA), as were the supply of payload shells, thermal design of the shells and, of course, many other essential aspects of the flight operations. Telemetry transmitters and other electronic components were provided by the Jet Propulsion Laboratory (J.P.L.), by Project Vanguard, and by the Army Signal Engineering and Development Laboratory.

Four flight payloads were built, fully calibrated and subjected to the full gamut of environmental tests. One payload was flown in Explorer IV, which entered a 51° inclination orbit successfully at 1506 UT on July 26, 1958, thus earning the astronomical designation Satellite 1958 Epsilon. The apparatus performed in all respects fully up to expectations, though the lifetime of its batteries was somewhat shorter than expected.

A second payload was flown on Explorer V, launched on August 24, 1958. This

rocket failed to go into orbit. However, the on-board apparatus functioned properly during its brief flight.

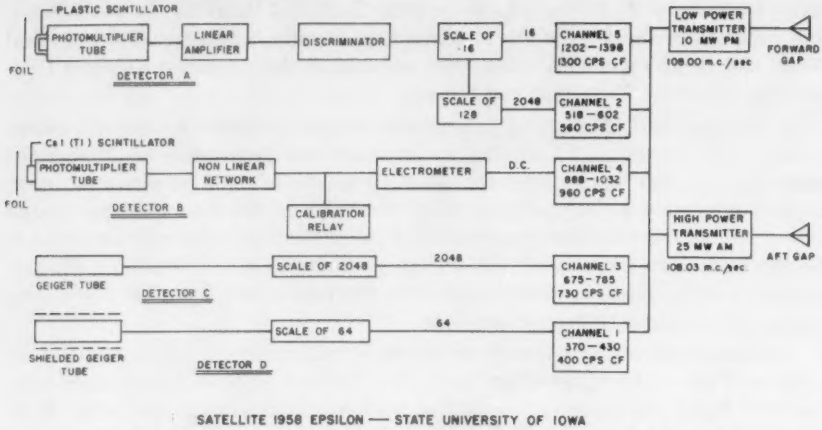

FIG. 1.—Block diagram of detectors and associated circuitry of Explorer IV.

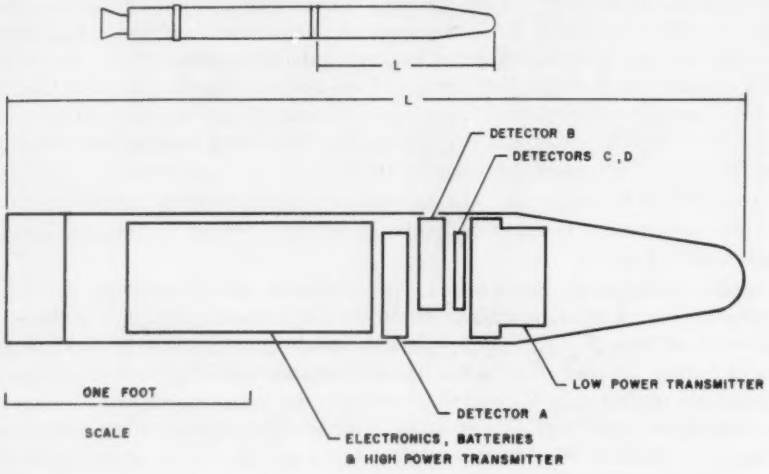

FIG. 2.—Outline sketch of the arrangement of principal components of the payload of Explorer IV (satellite 1958 epsilon); small diagram in upper part of the figure shows appearance of the orbiting body, of which the payload was a permanent part.

A full description of the observing equipment of Explorer IV and a preliminary account of the observations of the natural radiation have been published.[2] For convenience of the present reader Table 1 summarizes the principal characteristics of the apparatus. Figures 1 and 2 show the circuit schematic and physical arrangement respectively.

TABLE 1

SUMMARY OF DETECTOR CHARACTERISTICS (EXPLORER IV)

(See Reference 2 for more complete description)

Channel	Detector	Shielding	Sensitive to:	Geometric Factor
1	Geiger-Mueller Counter (Anton 302) Cylinder approximately 7 mm × 9 mm Scaler: 64	1.2 g/cm² Fe + 1.6 g/cm² Pb (minimum)	Electrons of $E > 5$ mev; protons of $E > 40$ mev; X rays of $E > 80$ kev with low efficiency	Omnidirectional geometric factor 0.14 cm² for minimum stopping power; 0.82 cm² for penetrability greater than 7 g/cm²
2	Plastic scintillator 0.75 cm diameter, 0.18 cm thick; Pulse Detector Scaler: 16	0.14 g/cm² of aluminum over window	Electrons of $E > 580$ kev; protons of $E > 10$ mev. Axis of detector ⊥ to payload axis X rays of $E > 300$ kev with low efficiency	0.040 cm² steradian; for minimum stopping power; 4.2 cm² steradian for penetrability greater than 5 g/cm²
3	Same as channel 1 except less shielding Scaler: 2048	1.2 g/cm² Fe (minimum)	Electrons of $E > 3$ mev; Protons of $E > 30$ mev; X rays of $E > 20$ kev with low efficiency	Omnidirectional geometric factor 0.14 cm² for min stopping power; 0.70 cm² for greater than 5 g/cm² stopping power
4	Cesium Iodide Crystal 0.76 cm diameter, 0.20 cm thick. Total energy-dc electrometer	1.0 mg/cm² of nickel and aluminum over window	Electrons of $E > 20$ kev; protons of $E > 400$ kev; x-rays Axis of detector — to payload axis	0.0235 cm² steradian for minimum stopping power 4.4 cm² steradian for penetrability greater than 5 g/cm²
5	Same as 2 except Scaler: 2048			

Reference 2 should be consulted for further experimental details as well as for discussion of interpretative aspects.

The Argus Detonations.—A series of three multi-stage rockets, fired from the *U.S.S. Norton Sound* in the south Atlantic Ocean, delivered small fission bombs to high altitude, where they were detonated. A variety of considerations led to the choice of site for detonations:

(*a*) An isolated firing area was desired for reasons of safety.

(*b*) An intermediate latitude was desired in order that the expected lifetime of the trapped particles be at least a few days in order to make possible comprehensive observations. At low latitudes the trapped particles spend a greater fraction of their paths in denser air. And at high latitude, the corresponding magnetic line of force goes farther from the earth in the equatorial plane and a reduction of trapped life time is to be expected due to perturbations in the outer reaches of the geomagnetic field.

(*c*) It was necessary that the chosen magnetic shell be such that it could be intercepted by a low-altitude satellite in an orbit inclined at not more than 51° to the equatorial plane (the highest inclination which was practical with existing U.S. facilities at that time). An intermediate latitude was preferable to a low latitude, also, because the satellite orbit would then intercept the shell of trapped particles at a less oblique angle.

(*d*) From the observations of Explorers I and III, it was known that the intensity of the natural radiation diminished rapidly, at a given altitude, as one receded

from the magnetic equator (either north or south). Hence the "background" radiation would be less at intermediate latitudes.

(*e*) In the south Atlantic, the scalar magnitude of the geomagnetic field is less than that at any other point on the earth at intermediate latitudes. Hence, for a given altitude of injection, the trapped lifetime of charged particles would be the greatest for this firing site (cf. reference 2).

(*f*) It was desirable that an island or other convenient observing site be located near the point conjugate to the firing point (the Azores Islands, in this case).

In Table 2 are summarized the pertinent data on the three Argus bursts.

TABLE 2

DATA ON BURSTS

	Argus I	Argus II	Argus III
Nominal yield	1 to 2 kilotons	1 to 2 kilotons	1 to 2 kilotons
Approximate time of burst	August 27, 1958, 0230 U.T.	August 30, 1958, 0320 U.T.	September 6, 1958, 2210 U.T.
Approximate geographic coordinates	38° S, 12° W	50° S, 8° W	50° S, 10° W

Nominal altitude of all bursts—480 km.

Nature and Scope of Satellite Observations.—During the month preceding Argus I (July 26 to August 26), an extensive body of observations served to establish the detailed spatial dependence of the natural, trapped radiation and many of its properties in terms of the characteristics of the four detectors which we had devised for the twofold program of observation.

Typical geometric relationships between the orbit of Explorer IV (Satellite 1958 Epsilon) and the chosen magnetic shell are illustrated in a schematic way in Figures 3 and 4. It is seen that in a favorable case the satellite orbit intersects the Argus shell four times during each revolution around the earth—at points A, B, C, and D. These intersections are, of course, at various altitudes, latitudes, and longitudes. In practice the number of observed intersections from which significant data were obtained was much less than four per orbit due to a variety of reasons:

(*a*) Many of the intersections were at such low altitudes that the intensity of Argus electrons was immeasurably small with the equipment used.

(*b*) In many cases, fewer than four intersections occurred due to the distortion and tilt of the chosen magnetic shell.

(*c*) Many intersections occurred beyond radio-telemetry range of any one of the receiving stations.

(*d*) Some of the intersections occurred at points where the artificially injected radiation was unobservedly small relative to the natural radiation. And, of course, this situation prevailed at all points within the reach of Explorer IV after the lapse of a sufficiently long time.

Table 3 gives a brief tabular summary of the numbers of observed intersections which yielded significant data. It should be noted that the batteries supplying the lower-powered of the two transmitters and telemetry channels 2 and 5 were exhausted on about September 3. The higher-powered transmitter and channels 1, 3, and 4 continued to operate properly until about the 21st of September. At this time the Argus III particles were still being detected above the background on occasional favorable intersections.

It appears probable that a residuum of the Argus electrons was also detected near the geomagnetic equator with Pioneer III on December 6, 1958. But no observable intensity was present there on March 3, 1959 (Pioneer IV).

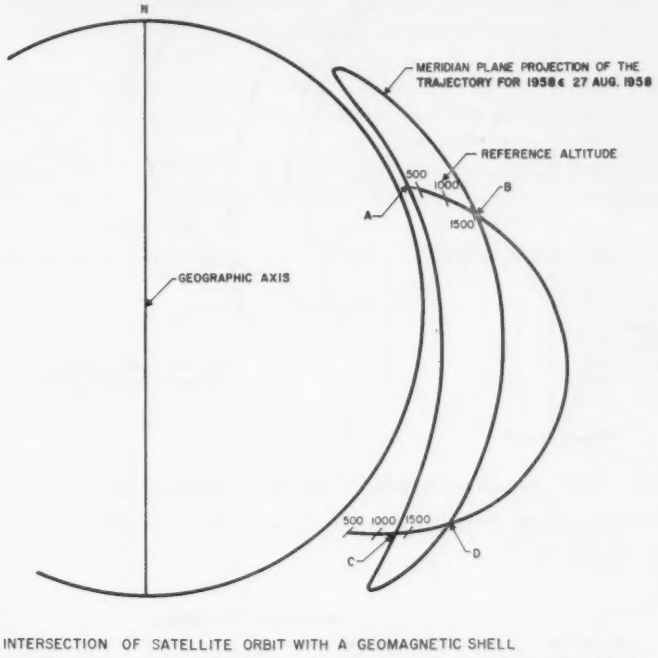

FIG. 3.—Illustrative diagram showing a sample geometric relationship between the orbit of satellite 1958 epsilon and a chosen magnetic shell at a given longitude.

Detailed Observations.—The "Argus effect" was easily and promptly observed by Explorer IV after each of the three detonations.

Figure 5 shows a record of measurements of the natural radiation taken on the day *before* Argus I. Figure 6 shows a record taken on a satellite pass through a similar geographic region about three and one half hours after Argus I. The same data are plotted with a linear scale of ordinates in Figure 7. The great peak which was intersected at 0608 U.T. on August 27 had no precedent in four weeks of previous observations of the natural radiation. Moreover, it was encountered on the first

TABLE 3

SUMMARY OF OBSERVED INTERSECTIONS GIVING SIGNIFICANT DATA

Burst	Number in Northern Hemisphere	Number in Southern Hemisphere	Period of Observations (1958)
Argus I	28	9	August 27 to September 6
Argus II	27	12	August 30 to September 6
Argus III	61	27	September 6 to September 21
Totals	116	48	.....

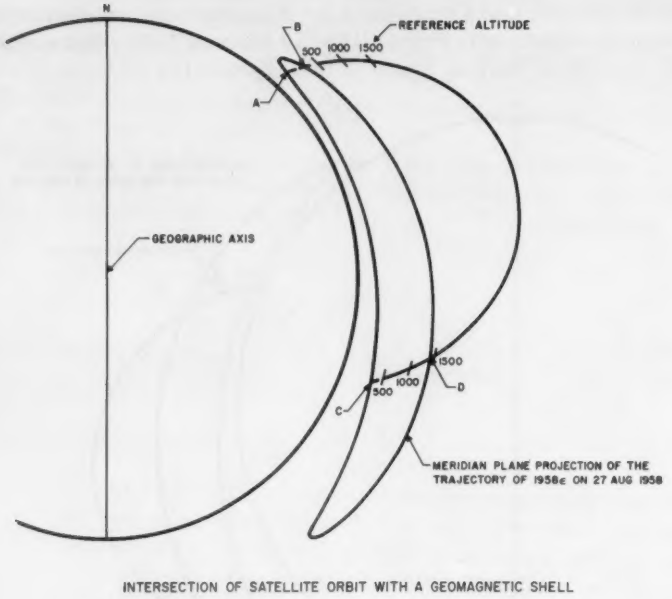

FIG. 4.—Same as Figure 3, except at a different longitude.

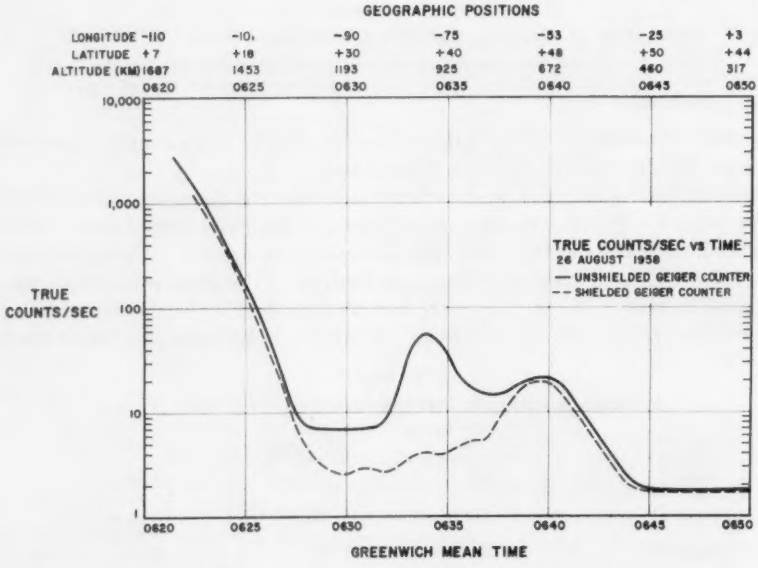

FIG. 5.—A plot of observations of the natural radiation on the day before Argus I.

observed intersection with the planned magnetic shell following the Argus I detonation. Further reasons for the validity of identifying such observed peaks with the "Argus effect" are as follows:

(*a*) The observed energy spectrum and nature of the radiation were found to be in essential agreement with that expected for the decay electrons from fission fragments.

(*b*) A peak with similar characteristics was found at every observed intersection of the orbit of the satellite with the appropriate magnetic shell, irrespective of latitude or longitude.

(*c*) The geometric thickness of the shell was similar to that of pretest estimates.

(*d*) The observed intensity of trapped electrons was in order-of-magnitude agreement with pretest estimates.

(*e*) The temporal decay of trapped intensity resembled pretest estimates.

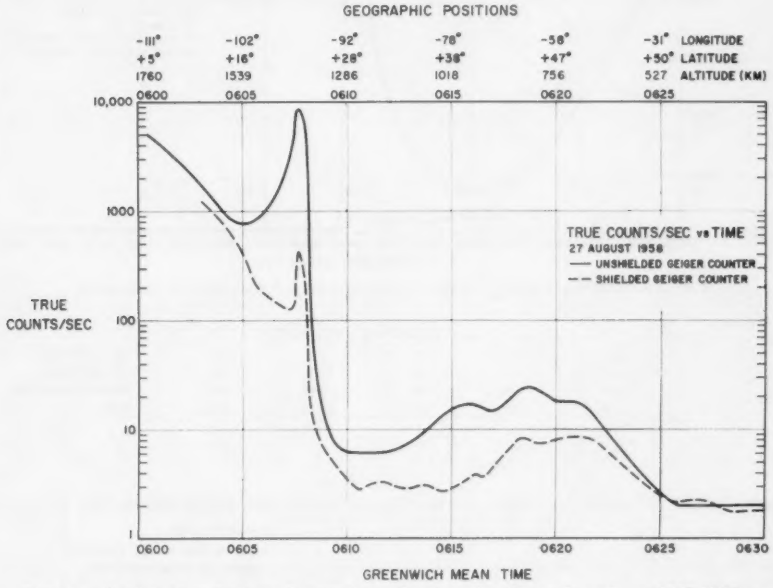

FIG. 6.—A plot of radiation observations showing the Argus I peak at 0608 U.T. on August 27, 1958, about three and one-half hours after the burst. The geographic region is similar to that of Figure 5.

The above remarks apply to the sequences of observations after each of the three shots—Argus I, II, and III. During the entire active period of Explorer IV (July 26 to September 21, 1958), no such peaks were observed except those having the proper temporal and spatial relationship to the artificial ones expected. The observing period included a number of important geophysical perturbations in the radiation around the earth. But these perturbations were of a quite different nature.

Figure 8 is a record on August 31, 1958, which shows the decaying peak due to Argus I (at 0446 U.T.) and the fresh peak due to Argus II (at 0449 U.T.); and Figure 9 is a sample record on September 9, 1958 which shows the peak due to Ar-

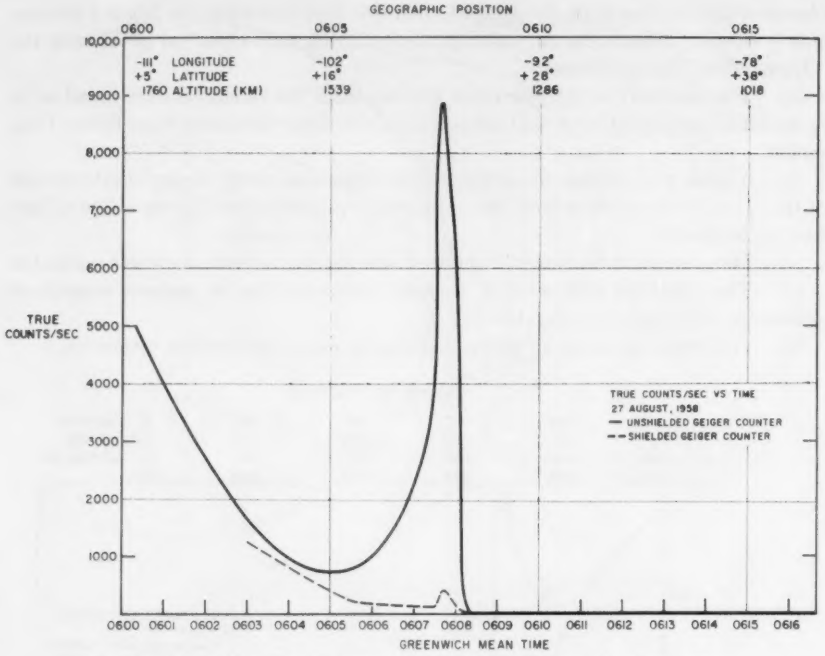

FIG. 7.—Same data as on Figure 6, plotted with a linear scale of ordinates.

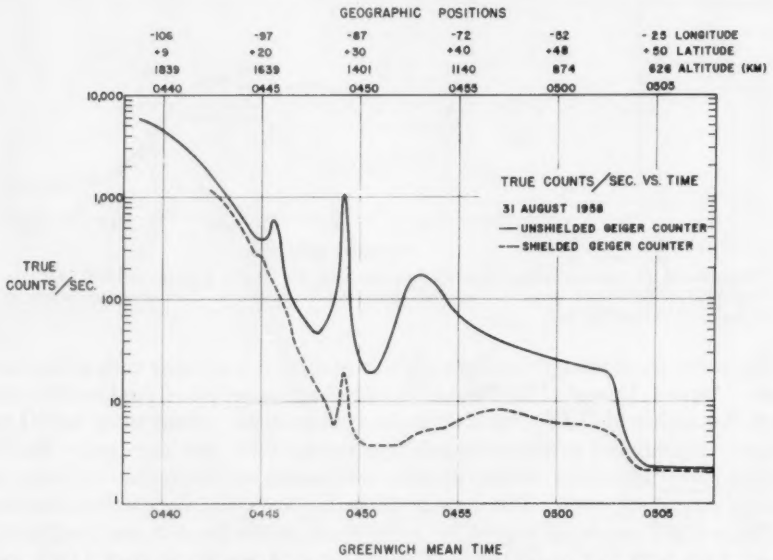

Fig. 8.—A plot of radiation observations on August 31, 1958 showing the decaying Argus I peak (at 0446 U.T.) and the fresh Argus II peak (at 0449 U.T.).

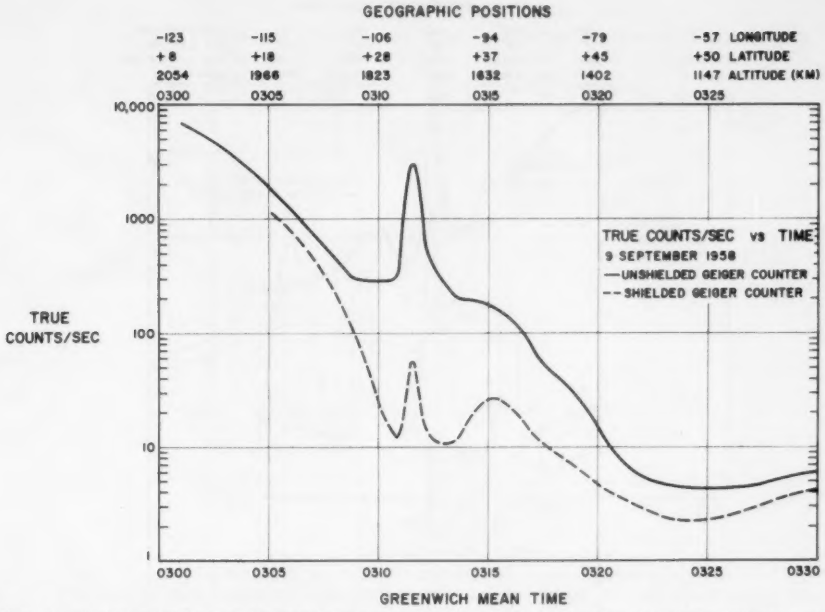

Fig. 9.—A plot of radiation observations on September 9, 1958, showing the Argus III peak at about 0312 U.T.

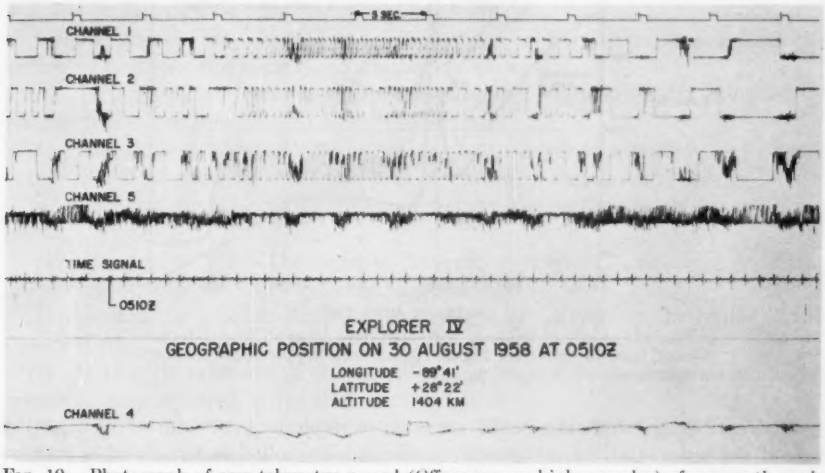

FIG. 10.—Photograph of raw telemetry record (Offner pen-and-ink recorder) of a pass through the Argus II shell on August 30, 1958.

gus III. Figure 10 is a photograph of a raw telemetry record of the passage of Explorer IV through the Argus II shell and Figure 11 gives plots of the reduced, corrected data from all four detectors. These two figures exemplify the data reduction procedure.

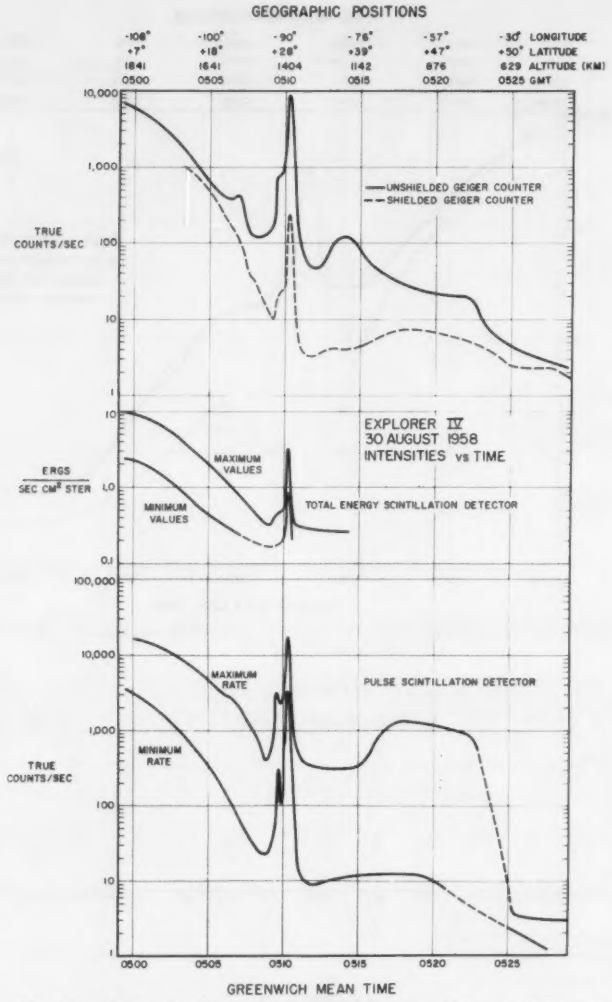

FIG. 11.—Plots of reduced, corrected data from all four detectors as derived from the record of Figure 10 and from the pre-flight calibrations of detectors.

Figure 12 (expanded time scale) gives an example of the variation of counting rate of the pulse scintillation detector as the satellite rolled and tumbled in its passage through the Argus I shell. As previously described[2] this type of "random" angular motion is very advantageous for the measurement of angular distributions. The angular distributions of the Argus electrons (as observed near their mirror altitudes) were found to be disklike in nature, as suggested by magnetic-trapping theory and as extensively observed on the natural radiation.

In each set of observations (cf. Figures 10 and 11), the output of each detector

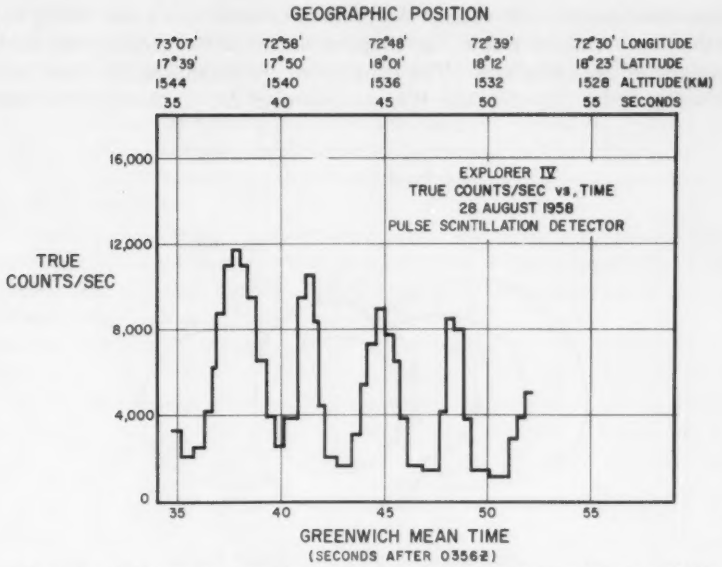

FIG. 12.—A plot on an expanded time scale of the detailed counting rate of the pulse scintillation counter as a function of time during passage through the Argus I shell. This plot illustrates the disk-like angular distribution of the radiation. The tumble period was 7 seconds.

was converted either to true counting rate (channels 1, 2, 3, and 5) or to absolute directional energy flux (channel 4) using preflight calibration data; and from the corrected plots the following quantities were estimated:

(*a*) Net intensity above background of observed Argus electrons, obtained by subtraction of the interpolated background values.

(*b*) The time width of the Argus peak at half-maximum intensity.

(*c*) The time width of the Argus peak at one-tenth maximum intensity.

(*d*) The local maximum and minimum values of intensity near the center of the peak.

(*e*) The time at which the center of the peak occurred. Then, from the ephemeris of the satellite[2] the following were tabulated:

(*f*) Geographic location (latitude, longitude, and altitude) of the center of the observed peak.

(*g*) Geometric thickness of the "shell" of electrons at half-maximum intensity, measured perpendicular to the shell.

Organization and Analysis of Observations.—The work sketched in the present section was considerably advanced by a ten-day study session arranged and conducted by the Department of Defense during February, 1959, at the Lawrence Radiation Laboratories in Livermore, California. The group specifically concerned with satellite observations consisted of the following persons: George Bing (Chairman), Donald Chandler, Rolf Dyce, William Karzas, John Killeen, Charles Lundquist, Hans Mark, Carl McIlwain, Paul Nakada, Theodore Northrup, Ralph Pennington, Russell Shelton, James Van Allen, and Ernest Vestine.

As mentioned earlier, the various observations occurred at a great variety of positions in space. And, of course, the irregular nature of the geomagnetic field produces essential complications. The foundations for organizing the data are indicated in Figure 13. By the basic Poincare-Stoermer-Alfven theory of the trapping

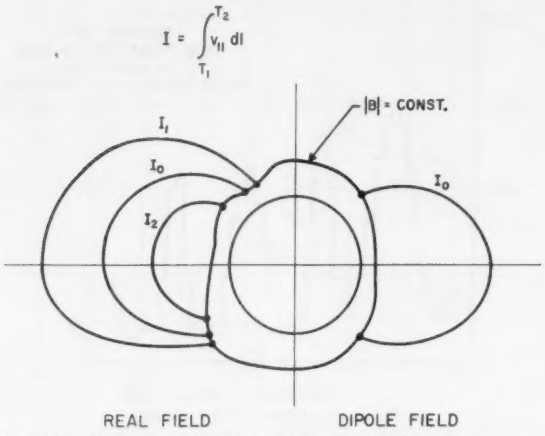

FIG. 13.—A diagram to illustrate the principles of conservation of μ and I in geomagnetic trapping. (See text.)

of charged particles in the geomagnetic field the magnetic moment of a spiraling particle is an adiabatic invariant of the motion. That is

$$\mu = \frac{{}^1/_2\, m\, v_\perp{}^2}{B} = \text{constant}$$

Since the speed v (having components $v_\parallel$ and $v_\perp$ at any moment parallel and perpendicular, respectively, to the magnetic vector B) is also a constant of the motion in a static magnetic field, the turning points (or mirror points) of the oscillatory, helical path of a given particle always occur at the same scalar value, B. Indeed, the locus of the turning points of all particles of given magnetic moment is a surface of B = constant (cf. Figure 13).

Also, as first conjectured by Rosenbluth and Longmire[3] and later discussed in detail by others, the line integral along a line of force between turning points for a given particle:

$$I = \int_{\tau_1}^{\tau_2} v_{||}\, dl = v \int_{\tau_1}^{\tau_2} \sqrt{1 - B/B_\tau}\, dl$$

is an adiabatic invariant (under an important class of physical conditions) of its motion, where B_τ is the scalar value of B at the turning points τ_1 and τ_2.

This latter principle makes possible the identification of a unique sequence of magnetic lines of force which constitute a single valued, three-dimensional surface on which the guiding center of the particle will forever lie—to the extent that the conditions for the conservation of μ and I are met—as it moves about in the irregular geomagnetic field. The argument is illustrated in Figure 13. Let the surface B = const. shown there represent the locus of turning points for a particle having a given

magnetic moment μ. Let the particle's motion, at a chosen period of time, be along the line of force shown in the right—hand side of the figure with the integral I having the value I_0. The question then is: Along which of the infinite number of segments of lines of force having values of I—I_0, I_1, I_2, etc. (sketched in the left-hand side of Figure 13)—and having turning points on the specified surface of constant B will the guiding center of the particle's trajectory be moving at some later time? The Rosenbluth principle assures one that it will be the segment characterized by I_0, or more precisely by I_0/v.

To the extent that the above discussion is applicable to electrons having energies up to several million electron volts and trapped in the real geomagnetic field it may be expected that an Argus shell of trapped electrons will maintain its integrity and will be fixed with respect to the earth.

Moreover, elementary consideration of perturbations by localized magnetic irregularities and by atmospheric scattering and energy-loss indicates that the tendency for diffusion of guiding centers transverse to the shell will be greatly exceeded by the tendency for diffusion of turning points along the shell and that the latter tendency will result in progressive leakage of particles downward into the denser atmosphere and eventual loss.

Hence, the temporal decay of the total number of trapped particles resulting from an impulsive injection may be expected to proceed by way of loss of particles out of the ends of the shell; and it may be further expected that the geometric thickness and position of the shell with respect to the earth will be substantially constant.

In the first instance, the observations following each detonation are functions of four apparently independent parameters—latitude, longitude, altitude, and elapsed time from the time of the detonation.

The considerations of the preceding paragraphs (which have been most clearly advocated by T. Northrup and implemented by E. H. Vestine) suggest that important progress can be made in systematizing the observations for the purpose of interpretation if the three positional parameters, latitude, longitude, and altitude, be supplanted by a single one—the scalar value B at the position of the observation. Apart from perturbations whose effects are important during one longitudinal precession period (order of a fraction of an hour in the present case), a surface of constant B may be expected, at a given time, to be a surface of constant particle intensity irrespective of the distorted shape of this surface in the real geomagnetic field.

Hence, it may be hoped that observable features of the trapped Argus radiation will be functions of only two parameters, B and elapsed time.

The geomagnetic field has a slow secular rate of change and B is not accurately known as a function of geographic coordinates during August—September, 1958. In the present work Vestine and associates have derived pertinent B values from the recent British 48-coefficient potential function (era of 1955). Small errors also arise from the imperfect knowledge of the orbit of the satellite (± 10 km) and from the observational uncertainty in locating the center of the shell of electrons.

Preliminary Summary of Significance of Observations.—(1) Thickness of Argus shells: The observed mean thickness of the Argus shells at half-intensity, measured perpendicular to the boundaries of the shells, was about 90 km for events I and II and 150 km for event III. The thickness of the shells did *not* have a systematic dependence on elapsed time, but the scatter of the measured values of thickness is of the

order of 30 per cent and no lesser broadening (or thinning) of the shell during the observing period could be discerned. It is presumed that the basic shell thickness was determined by the very complex injection situation at and immediately after the detonation in each case. This surmise is supported by crude consideration of the dynamics of the detonation in the earth's magnetic field.

It does seem to be of importance that the leakage of particles from the shell (by energy loss and "longitudinal" diffusion) proceeded at a sufficiently rapid rate that the transverse diffusion was, in general, unobservably small. (See, however, a later section remarking on the influence of the geomagnetic storm of September 3 ff.) And it may be noted that the magnetic shell in question reaches outward to a radial distance of about two earth radii from the center of the earth.

These results, of course, provide a quantitative validation of the principle of conservation of I in the geomagnetic field.

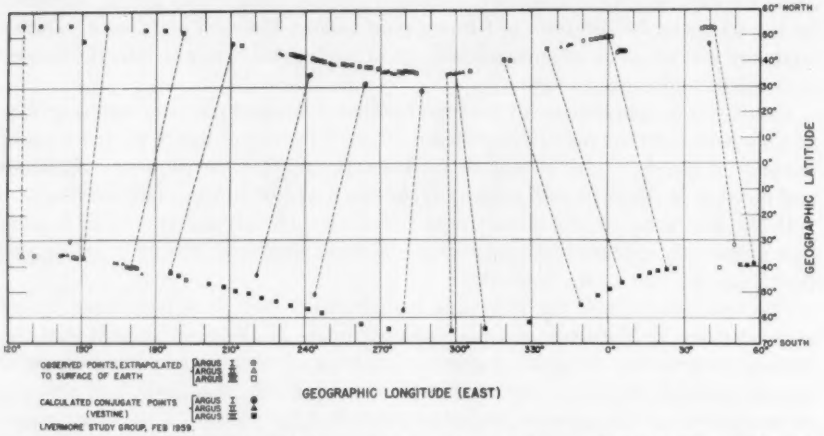

FIG. 14.—Plot of (extrapolated) observed intersections of Argus shells" I, II and III with surface of the earth, and computed conjugate points (after Vestine and Pennington). (Livermore study group.)

(2) *Position of the shell in space:* Each observation of the center of each Argus shell (I, II, and III) was projected down to the surface of the solid earth by use of a simplified model of the geomagnetic field, the eccentric dipole model. A comprehensive plot of the results of this process is shown in Figure 14, together with conjugate points calculated with the full 48-coefficient potential function. The three resulting loci of points, interlaced in time, are notably smooth and regular suggesting the positional stability of the several shells during the observing period. A more critical examination of this question can be had by examination of Figure 15, a large scale plot of projected observations on Argus III. In view of the interlacing of observations at various sequences of times, we conclude that the shell could not have drifted monotonically in latitude by more than about 0.03° of latitude per day (about 2 km of distance per day at 1500 km) without having resulted in an unreasonable "ripple" in the implied geomagnetic field contour.

These results are in the nature of a confirmation both qualitatively and quantitatively of the principles which have been sketched in preceding sections.

(3) *Angular distribution:* As typified by Figure 12 (see also reference 2), the angular distribution of the trapped electrons was always observed to be disklike in character with an angular thickness of the order of $\pm 20°$ or less. In terms of the basic dynamics of the trapping of charged particles this fact shows that, at the relatively low altitudes of observation, most particles being measured have their turning points not more than a few hundred kilometers below the observing point —a quite reasonable result.

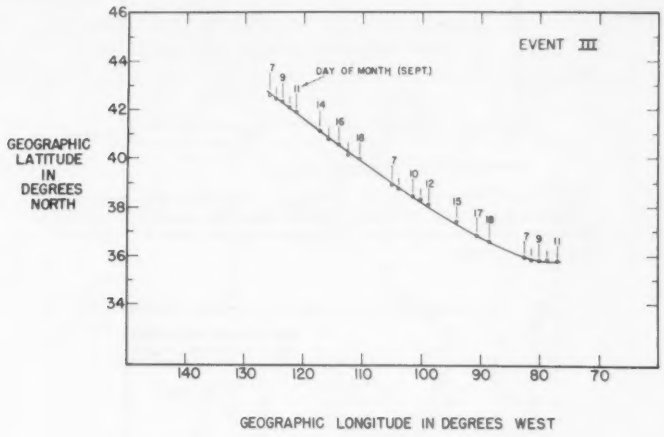

FIG. 15.—An expanded plot of a portion of Figure 14.

(4) *Trapped lifetimes:* The product of the maximum observed intensity, C, at the center of the shell times the thickness, W, of the shell at half-intensity has been taken as a measure of the total intensity of trapped particles (within the detection range of the detector in question). Several preliminary plots of this quantity, CW, versus elapsed time, t, for two different ranges of B are given in Figures 16, 17 and 18. The plots are of data from channel 3 (the unshielded Geiger tube). Closer examination of several selected cases suggests that the scatter of data will be reduced by taking proper account of the average aspect of the satellite equipment as it passed through the shells. It is seen that CW varies approximately as the negative first power of the elapsed time. This law of time-decay of trapped intensity, namely $CWt = \text{const.}$ is in general accord with a theory of the loss mechanism due to Ernest C. Ray of this laboratory (to be published). The above curves based on our channel 3 data (the most extensive and simple body of data) pertain to an effective electron energy of about 3.5 mev. By comparison with the theory Ray finds that the effective atmospheric density at 1,000 km required to reduce the intensity as rapidly as observed is at least 10 times as great as one would suppose from the plausible extrapolation of the Jastrow "satellite-drag" atmosphere.

It is of interest to note that the Argus shells occurred (Fig. 19) within the "slot"

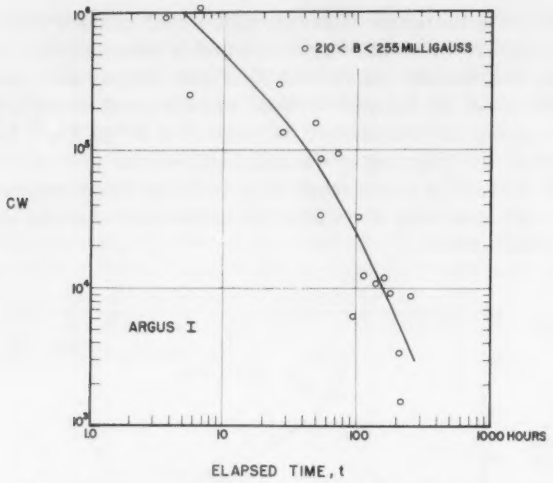

FIG. 16.—The product of maximum true counting rate, C, of channel 3 at center of Argus I shell times the geometric width (thickness) at half-maximum, W, versus elapsed time, t. (Livermore Study Group.)

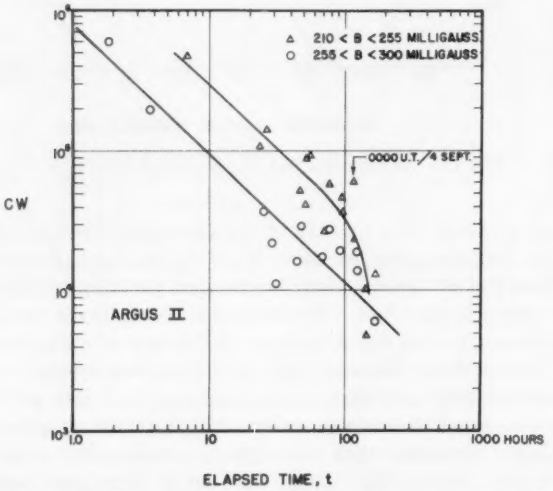

FIG. 17.—The product of maximum true counting rate, C, of channel 3 at center of Argus II shell times the geometric width (thickness) at half-maximum, W, versus elapsed time, t. (Livermore Study Group.)

between the two natural radiation zones.[4, 5] It may be that the atmospheric density is anomalously high there.

An alternative line of speculation would suppose that the rapid decay was, at least in part, due to "magnetic diffusion" due to small scale and perhaps time-de-

pendent irregularities of the geomagnetic field. A more conclusive discussion of this point is in process, making use of the results from all four detectors (different effective particle energies).

The gross decay rate is presumably a composite of atmospheric and magnetic effects.

The present experimental results on artificially injected electrons are of far reaching importance in understanding the dynamics of the natural radiation.

The discontinuously rapid drop of intensity of Argus I and Argus II on about the 4th of September suggests that the great magnetic storm of that date may have caused a greatly enhanced rate of loss of stored particles. Also, the thickness of the Argus II shell appears to have increased significantly during the same period.

These observations may provide an important clue to the understanding of the structure and intensity of the natural radiation belts.

(5) *Injection efficiency:* By means of our Explorer IV observations, it is possible

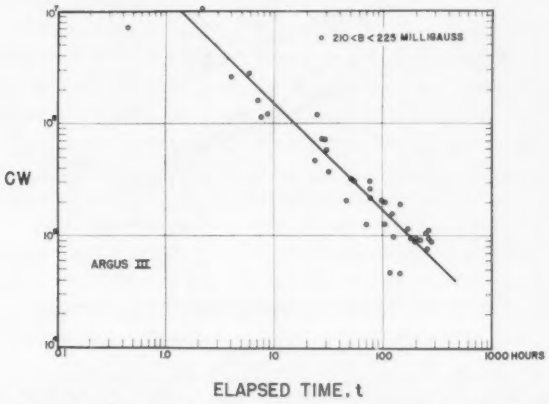

FIG. 18.—The product of maximum true counting rate, C, of channel 3 at center of Argus III shell times the geometric width (thickness) at half-maximum, W, versus elapsed time, t. (Livermore Study Group.)

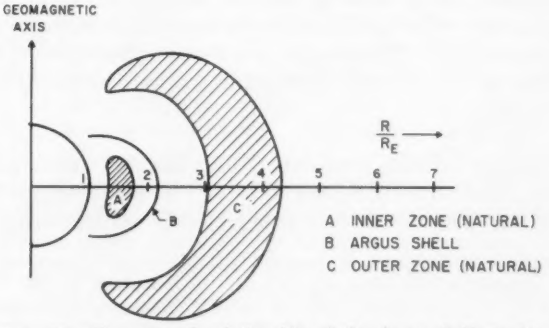

FIG. 19.—The general relationship of the Argus shells to the structure of the natural radiation zones. See references 4 and 5.

to estimate the total number of trapped Argus electrons having turning points below the highest altitude of observation. This estimate, made as of zero elapsed time, serves to place a lower limit on the fraction of electrons effectively injected into the geomagnetic field, since the nominal yield and nominal spectrum of electrons are known.

(6) *Distribution of turning points:* Figure 20 is a plot of CWt versus B for Argus I, II, and III. The distribution of turning points along a line of force is implicit in these curves. It is seen that there is a monotonic increase in the density of turning points with decreasing B (increasing altitude). It is therefore evident that injection of the longer-lived electrons occurred, in large part, at altitudes much greater than that of the atomic detonations.

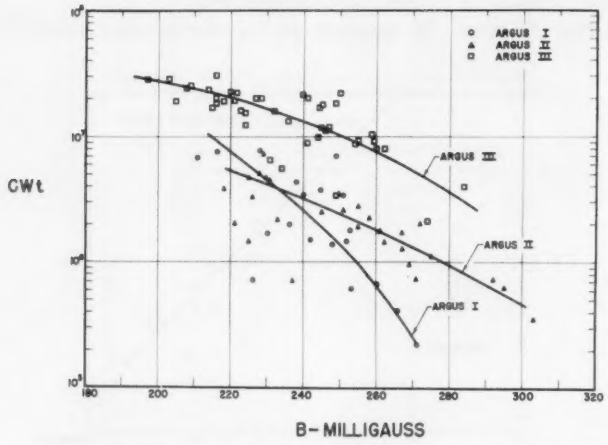

FIG. 20.—The triple product of C times W times t for channel 3 versus scaler magnetic field intensity B for Argus I, Argus II and Argus III. (Livermore Study Group.)

(7) *The general geomagnetic field:* The geometric form of the Argus shells is known to a precision of the order of ± 10 km with respect to the earth. Hence, these experiments provide a significant new method for the harmonic analysis of the general geomagnetic field. This analysis is not yet completed but it is already clear that at least the dipole and quadrupole terms will be determined with useful accuracy.

The Argus work at the State University of Iowa has been supported, in large part, by contract no. DA-11-022-ORD-2788 with the Chicago Ordnance District of the Department of the Army, which support is gratefully acknowledged.

[1] Van Allen, J. A., G. H. Ludwig, E. C. Ray, and C. E. McIlwain," Observation of High Intensity Radiation by Satellites 1958 Alpha and Gamma," *Jet Propulsion*, **28**, 588–592 (1958).

[2] Van Allen, J. A., C. E. McIlwain, and G. H. Ludwig, "Radiation Observations with Satellite 1958ϵ," *Journal of Geophysical Research*, **64**, 271–286 (1959).

[3] Rosenbluth, M. N., and C. L. Longmire, "Stability of Plasmas Confined by Magnetic Fields,"

Annals of Physics, **1**, 120–140 (1957). (Apparently the first paper on the matter in the open literature.)

[4] Van Allen, J. A., and L. A. Frank, "Radiation Around the Earth to a Radial Distance of 107,400 km," *Nature*, **183**, 430–434 (1959).

[5] Vernov, S. N., A. Ye. Chudakov, P. V. Vakulov, and Yu. I. Logachev, "Study of Terrestrial Corpuscular Radiation and Cosmic Rays During Flight of the Cosmic Rocket," *Doklady Akademiia Nauk S.S.S.R.*, **125**, 304–307 (1959).

PROJECT JASON MEASUREMENT OF TRAPPED ELECTRONS FROM A NUCLEAR DEVICE BY SOUNDING ROCKETS

BY LEW ALLEN, JR., JAMES L. BEAVERS II, WILLIAM A. WHITAKER, JASPER A. WELCH, JR., AND RODDY B. WALTON

PHYSICS DIVISION, RESEARCH DIRECTORATE, AIR FORCE SPECIAL WEAPONS CENTER, AIR RESEARCH AND DEVELOPMENT COMMAND, KIRTLAND AIR FORCE BASE, NEW MEXICO

Equipment.— Project Jason is the name for the Air Force Special Weapons Center's participation in the Argus experiment. It consisted of the firing of 19 high-altitude sounding rockets to measure the electrons created by the Argus detonations. The carrier vehicle used was a five-stage solid propellant rocket consisting of an Honest John for the first stage, Nike boosters for the second and third stages, a Recruit for the fourth stage, and a T-55 for the fifth stage. These were capable of delivering a 50-pound payload to an altitude of 800 km when launched at an elevation of 80°. They were launched from three sites: Cape Canaveral, Florida (Air Force Missile Test Center); Wallops Island, Virginia (NASA Pilotless Aircraft Test Station); and Ramey Air Force Base, Puerto Rico. Table 1 gives some particulars of each launching. The flights are referred to by Patrick Air Force Base test number and are listed in chronological, rather than numerical, order. The table shows the launch site, date of launch, launch time after the appropriate burst, and the apogee and splash coordinates of the flight. Also shown are the rocket spin and tumble periods.

The instrumentation used in this project was basically a radiation sensing system composed of eight Geiger-Müller tubes and a system for providing a data link to ground receiving. Inasmuch as new phenomena were being investigated, it was necessary to provide a radiation detecting system with various thresholds and dynamic ranges that would best satisfy the expected conditions. Several tubes were collimated to observe the angular distribution of the electron flux. The selected values for thresholds and dynamic ranges proved quite adequate. In the design of the instrumentation system, effort was made to provide a system with a minimum number of components and maximum reliability. The package was designed and constructed under contract by Lockheed MSD; Figure 1 shows a breakdown of the complete instrumentation package. The transducer head, some aspects of which are shown schematically in Figure 2, consisted of eight G-M tubes arranged around the circumference of the forward portion of the instrumentation package. The tubes were protected during the lower atmospheric portion of the flight by a nose cap which was jettisoned at an altitude of 400,000 feet, approximately. After jettison of nose cap, all the detector circuitry became

TABLE 1

TEST DATA

Launch coordinates for the Jason Flights were: Wallops (75° 29′ W, 37° 50′ N); Patrick (80° 32′ W, 28° 27′ N); Ramey (67° 7′ W, 18° 32′ N)

Event/ Flight	Launch Site	Date	After Burst (Launch Time), hr:min	Perform-ance	Apogee, km	Splash Latitude	Splash Longitude	Spin Period, sec	Tumble Period, sec
Pacific I	Johnston I	1 Aug	..	..	..	..	..	..	..
Pacific II	Johnston I	12 Aug	..	..	..	..	..	..	..
1822	Patrick	15 Aug	65:46	OK	693	27.91° N	76°49′ W	2.0	2 25
1841	Ramey	20 Aug	182:57	Failure	..	..	..	..	..
1859	Wallops	25 Aug	319:43	Failure	..	..	..	..	..
Event 1	..	27 Aug	..			..	..	..	..
1909	Patrick	27 Aug	1:03	OK	937	30.62	74.01	15.1	120
1914	Ramey	27 Aug	1:54	Failure	..	..	..	..	..
1917	Ramey	27 Aug	4:12	OK	817	25.40	68.49	2.33	2.2
1913	Wallops	27 Aug	4:59	Failure	..	..	..	..	..
Event 2		30 Aug	..	..	..	..	..	..	..
2019	Wallops	30 Aug	0:28	OK	817	29.90	74.88	5.0	3.80
2022	Patrick	30 Aug	1:11	OK	878	28.19	70.21	1.07	38.5
2021	Wallops	30 Aug	1:58	OK	830	31.51	80.30	6.5	30.0
2023	Ramey	30 Aug	2:32	OK	825	25.51	70.86	3.8	8.8
2025	Patrick	30 Aug	3:16	OK	699	27.32	70.52	3.2	1.20
2024	Wallops	30 Aug	4:01	OK	815	27.56	72.56	1.97	17.4
2027	Wallops	30 Aug	18:42	OK	745	30.08	71.08	1.50	1.80
2026	Ramey	30 Aug	19:43	Failure	..	..	..	..	..
2020	Patrick	31 Aug	20:47	OK	800	25.96	67.47	5.4	13.2
2041	Ramey	2 Sep	87:42	Failure	..	..	..	..	..
2042	Wallops	2 Sep	88:43	OK	789	29.61	70.19	4.3	9.9
2043	Patrick	2 Sep	90:55	OK	789	27.96	67.27	20.7	138.5

exposed to the radiation environment. The possibility of arcing, necessitated "plotting" of all electrical connections in the instrumentation section. Table 2 gives specific information on each tube used in the transducer head.

The output pulses of the Geiger-Müller tubes were sequentially sampled and transmitted directly to ground. Calibration as of the FM deviation was not necessary, as no analog data were involved. The telemetry system was AM/FM

TABLE 2

DETECTOR DETAILS

Channel	Window	Direc-tional	Area, cm^2	Material	Window Thickness, mg/cm^2	Approx. Max. Flux	Min. Energy, kev
1	Side	Yes	13.	Steel	30	1.3×10^4	190
2	End	Yes	0.08	Aluminum	28	5×10^6	170
3	Side	No	13.	Aluminum	400	3.3×10^3	1000
4	End	Yes	0.08	Brass	400	5×10^6	900
5	Side	Yes	13.	Steel	30	1.3×10^4	190
6	Side	No	18.	Brass	2000	1.7×10^3	4000
7	Side	No	13.	Aluminum	150	2.4×10^3	460
8	End	Yes	0.08	Aluminum	150	5×10^6	470

at 217.5 megacycles. Power output was a minimum of 6 watts. The radiating antenna was formed by the outer shell of the instrument package and the fifth stage rocket, each forming half of a resonant antenna system. The Air Force Missile Test Center range telemetry receiving equipment was used, particularly the TLM-18 60-foot parabolic dishes. The average received signal of all flights at apogee was 100 microvolts input to the receiver from the TLM 18.

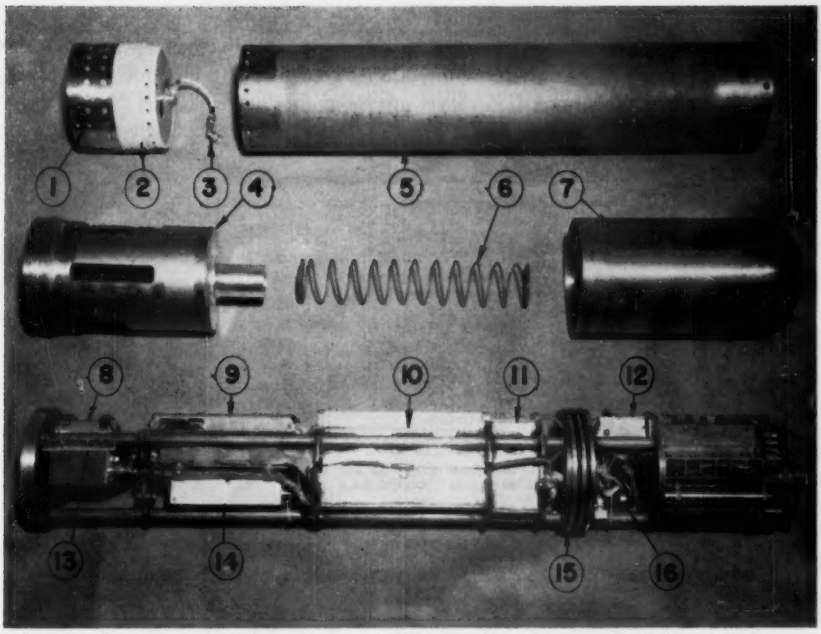

FIG. 1.—Instrumentation Components. (1) Package-to-rocket mating fixture, (2) Antenna insulator, (3) Antenna feed coaxial line, (4) Nose cap ejector mechanism housing, (5) Outer skin of instrument package, (6) Nose cap ejection spring, (7) Nose cap, (8) Acceleration actuated switch, (9) Commutator and modulation amplifier tray, (10) Plate supply battery box, (11) Filament supply battery box, (12) Timer and telemetry actuation switches, (13) RF power amplifier, (14) FM transmitter, (15) "O" ring seal, (16) Nose cap ejection timer.

Counters.—The relationship between counting rates and incident flux is given by

$$R = \int_{\Omega,E} G(E,\Omega)F(E,\Omega)dEd\Omega$$

where

R = counts per second,
G = absolute geometrical factor for electron of energy E incident within solid angle Ω with respect to the counter (cm^2 ster),
F = flux in electron/cm^2 ster Mev at energy E within Ω.

As will be shown, the electron flux is essentially confined to a plane perpendicular to the geomagnetic field; therefore, we may consider two simplifying cases: A, where the long axis of counters 1, 3, 5, 6, and 7 is perpendicular to the plane: and B, where the axis is parallel. For these cases the integration over the incident directions can be performed and the result expressed as:

$$R = \int_E G_0(E)F_0(E)dE$$

We note now that $F_0(E)$ = planar electron flux (electrons per cm^2 sec), i.e., a planar monoenergetic flux of F_0, would produce a counting rate of F_0 in an iso-

tropic detector having a transmission of unity and a cross sectional area of 1 cm^2. An isotropic flux F_0 would also count F_0 in the same detector, hence, G_0 is numerically the same as the absolute omnidirectional geometrical factor for these cases. Representative angular resolution curves are shown in Figure 5.

$G_0(E)$ is plotted in Figures 3 and 4. The dashed curves in Figure 3 represent

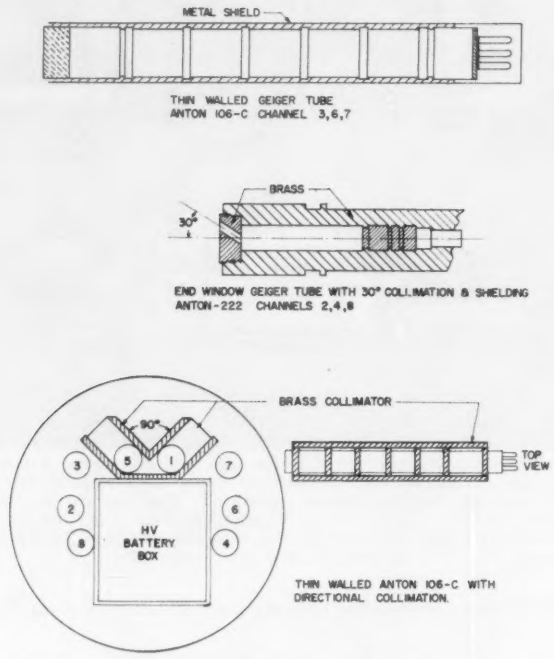

FIG. 2.—Schematic representations of the Geiger tubes and their orientation in the instrument head.

the results for orientation "A" when these differ from those calculated for orientation "B." In these cases the results for orientation "B" are indicated by solid curves. Because of the narrow collimations used for the small end windows 2, 4, and 8, and the side detectors 1 and 5, the transmissions of these detectors were the same for both orientations considered. The calculation for detector 6 was treated somewhat uniquely because the thickness of the brass shield, 2 gm/cm^2, was so great that the root mean square scattering angle of a 4 Mev electron transversing the shield would exceed $\pi/2$. In this case, an electron penetrating the shield would not remember its direction of entry into the shield; hence, the transmission was a function only of energy and shield thickness.

Dead time corrections were made to all data assuming a value of 10^{-4} seconds, the value recommended by the tube manufacturer. Experiments were performed which verified this value for a representative number of tubes used in this project.

Data Requirements and Data Reduction.—The exact data required in this project

were counting rates as a function of time for each G-M tube, and the rocket trajectory. Continuous recording of each channel was not accomplished. A commutator sampled each channel five times per second for a period of $^1/_{75}$ of a second.

Because of the low frequency response of the tape recorder, synchronization and reference pulses suffered severe differentiation; however, this had no effect on the quality of the counter pulses recorded. To provide synchronization and

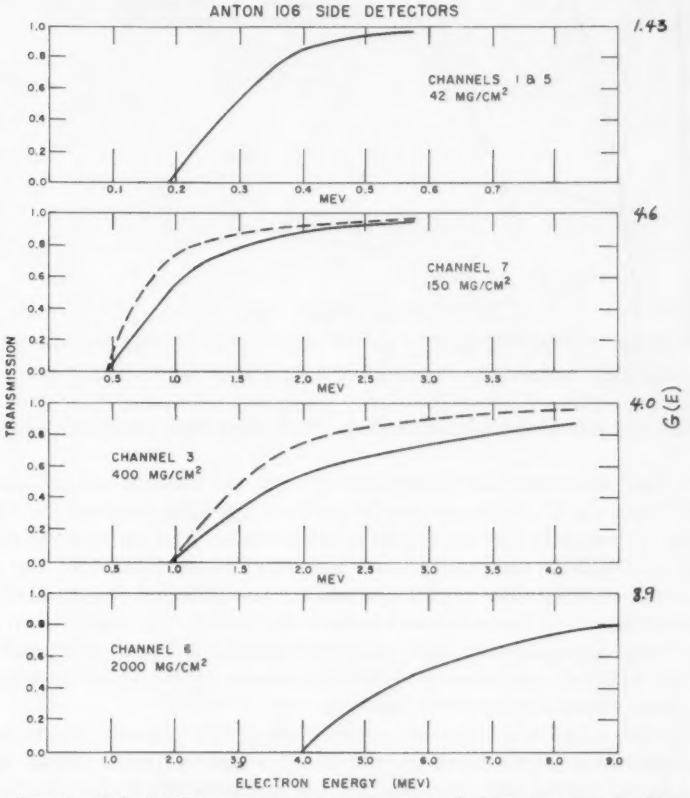

FIG. 3.—Calculated transmissions as a function of electron energy for high sensitivity detectors. The dashed curves refer to orientation "A" and the solid curves to orientation "B."

reference pulses for the data reduction purposes, the composite signal was used to modulate a 22 kc sub-carrier oscillator at the ground stations. The discrimination of this sub-carrier returned the desired information. Playback of the data was made at low speed onto an oscillograph with a three kilocycle response, and individual pulses were counted on the resulting record.

The total missile flight time for the 13 firings of interest was about 200 minutes. It was found sufficient to analyze only about one-half of the total flight time and this playback required 21 miles of oscillograph paper. Data processed in this manner

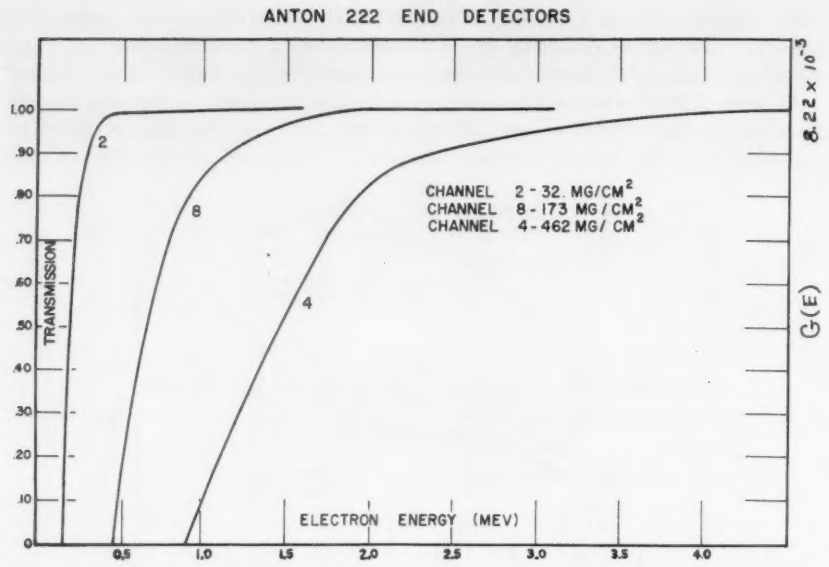

FIG. 4.—Calculated transmissions of low sensitivity and detectors as a function of electron energy.

were carefully edited to minimize reading errors which were always less than 10 per cent.

Each flight trajectory was determined by using the azimuth and elevation time histories from the TLM-18 antennas located at Cape Canaveral and the Island of Antigua. These antennas are 60-foot parabolic dishes used to track the data link telemetry. Ballistic trajectories were calculated, which minimized the squares of the errors between these trajectories and the azimuths and elevations observed. A summary plot of all the trajectories is shown in Figure 6. In order to correlate the results from all rockets with their trajectory variations, launch point and launch time, any longitude variations have been suppressed by projecting all trajectories onto a plane containing the 75° W meridian.

It was also desirable to construct a model of the earth's magnetic field in the plane of the projection. Since the observed field differs greatly from the dipole approximation in this region, the observed dip angle at the earth's surface is used. For the magnetic lines of force, arcs of circles were drawn which had the observed dip at the surface and a curvature derived from the dipole field. The lines were labeled by the geographic latitude at which they intersect the surface, and thus form a magnetic field line coordinate.

In Figure 6 the projected trajectories are shown together with the constructed magnetic lines. It should be noted that the geometry of this plot is distorted. First, the curvature of the earth is suppressed; second, the abscissa is in degrees of latitude which are about 105 km while the ordinate is given in hundreds of kilometers. This distortion causes the magnetic field lines to appear straight. It will become clear that use of the magnetic line coordinate was more natural and meaningful than was the use of geographic latitude.

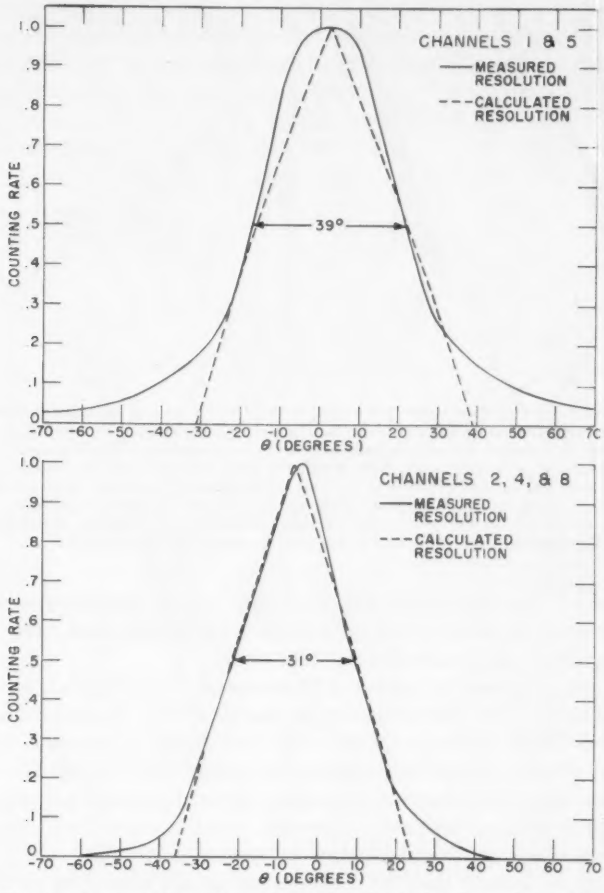

FIG. 5.—A comparison of experimental and calculated angular resolutions.

Should the dipole approximation be accurate, the geomagnetic latitude at the surface would be given by the latitude of the line of force plus eleven degrees. However, any comparison of this experiment with others at far distant places should be made on the basis of exact geographic position and height and a more accurate model of the earth's magnetic field.

Results.—As Table 1 indicates, there were 13 successful rocket flights out of a total of 19 attempts. One of these (1822) was launched prior to any of the South Atlantic shots and was therefore used as a background measurement. Actually, it was less than three days after a high altitude shot in the Pacific, and there are indications that this flight counted trapped betas from the decay of neutrons. Nevertheless, this flight served as a background measurement for the detectors having a threshold above the end point of the neutron beta decay spectrum. Of the other 12 flights, 2 were used on the first of the South Atlantic events, and 10

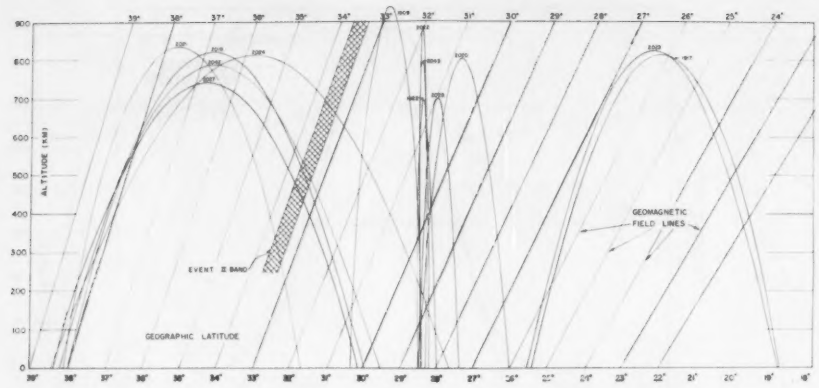

FIG. 6.—Jason rocket trajectories presented as altitude versus geographic latitude. Trajectories have been projected on the plane containing the 75° meridian. The position and width of the Event 2 band, together with the magnetic field lines are shown. The apparent launch points of the trajectories are not coincident, since a ballistic path was fitted to the unpowered portion of the flight. Slight angular differences at the end of the powered portion cause significant differences in apparent origins. On the ascending portion of the Wallops trajectories, the error is about plus or minus 10 km due to the angle of sight from Patrick. On the descending portion of the Wallops flight and for the other flights, the error is about plus or minus 2 km.

on the second. The third South Atlantic event was not monitored in this project. This report presents results from the second event for the most part, and unless otherwise specified, all remarks refer to this event.

Event 2 was monitored by a total of 10 successful rocket flights that covered the time period from H plus 28 minutes to H plus 91 hours. A well defined band of electrons was found immediately after the shot along a magnetic line of force which intersects the earth at a geographic latitude of 33.5° N, 75° W. This band persisted throughout the observation period and remained in a fixed position within the accuracy of the measurement. Increased counting rates (10 to 100 times back ground) were observed several hundred kilometers north and south of the band. In this report we assume that the observed counts are caused by electrons; however, the counts could conceivably be caused by any charged particle with the same range in the absorbers.

Results are given for such parameters as the position of the trapped electron band, the energy spectrum, and the rate of decay of the trapped particles. A discussion is made of how the orientation of the instrument package was determined unambiguously for several flights, and how it was determined from this that the electron flux is perpendicular to the magnetic field. Generally, the most useful channel for qualitatively following the events is channel 3, since its threshold was above the neutron decay beta spectrum and the observed low energy background. Furthermore, the threshold was still low enough to count a large fraction of the fission-product electrons.

Angular Distribution of Trapped Electrons.—The directionality of the radiation field was obtained from plots such as shown in Figure 7. In this figure, counts per commutator sample have been plotted against time (uncorrected for the dead time of the counters) for channels 1, 2, and 5 of Flight 1909. Zero time is an arbitrarily

chosen reference point. These detectors all have threshold of about 180 Kev; 1 and 5 are collimated at right angles to the missile axis and to each other, while 2 views the field at 30° to the missile axis. The counting data were extracted from flight 1909 because the spin and tumble periods and the orientation of the total angular momentum with respect to the magnetic field for this flight were optimum for showing the planar character of the electron flux.

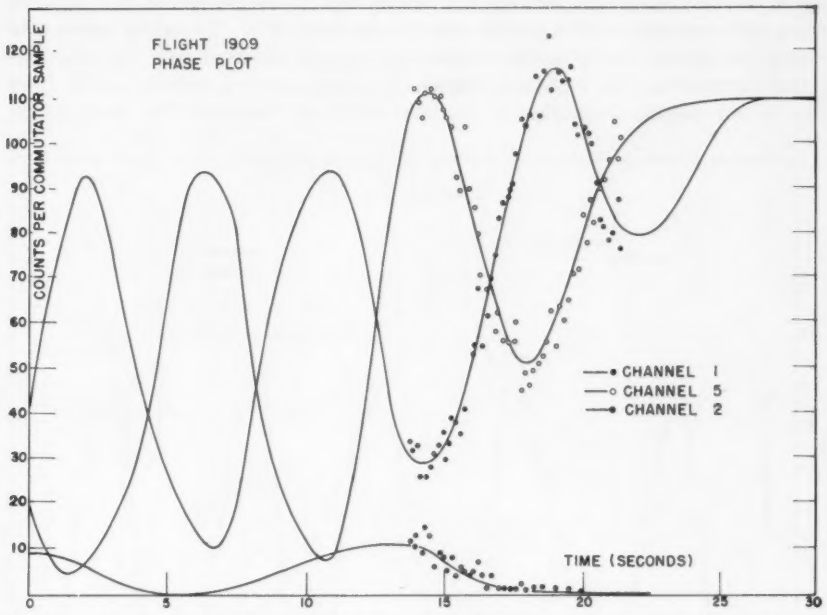

FIG. 7.—A phase plot for flight 1909. Both smoothed data and several typical points are shown.

It can be seen that the counts of channels 1 and 5 oscillate out of phase for the first 20 seconds and then coalesce. This behavior indicates that these counters at first pass in and out of the electron field as the rocket spins and then count at equal rates later in the tumble period when they are both completely in the field for several spin cycles. Thus, in the time period shown in Figure 7, the missile tumbles from a position where it is parallel to the plane of the radiation field (or perpendicular to the magnetic field) to a position where the axis is perpendicular to the plane of radiation. Because detectors 1 and 5 were almost completely saturated for counts greater than 100 per commutator sample, the peaks for these detectors are greatly depressed.

The angular distribution of the electrons was obtained using the known angular resolution of the collimated detectors. Flight 2019 was used in this study. This flight was launched from Wallops Island 28 minutes after event 2 and it passed through the electron band at H plus 41 minutes. It turned out to be a particularly fortuitous flight for this measurement because of the orientation of the rocket in space and its spin period. The analysis was done for two portions of the flight, in the wings and in the band.

Figure 8 shows the counting rate observed with channels 1 and 5 at an altitude of 465 km and along the 38° magnetic line. At this position, the package is 5° (of magnetic line) North of the electron band. Dead time corrections have been applied to the data. The uncertainties shown in the figure are statistical; errors from dead time corrections are negligible. The counting rates of the other directional counters were too low to be statistically meaningful. As indicated in Figure 8, the spin rate of the package was determined by measuring the time between consecutive maxima in the counting rate of detectors 1 or 5. Using the known spin rate it was possible to replot the data as counting rate versus angle. This manipulation showed that the measured angular distribution was indistinguishable from the known angular resolution of the detectors and therefore the electron flux

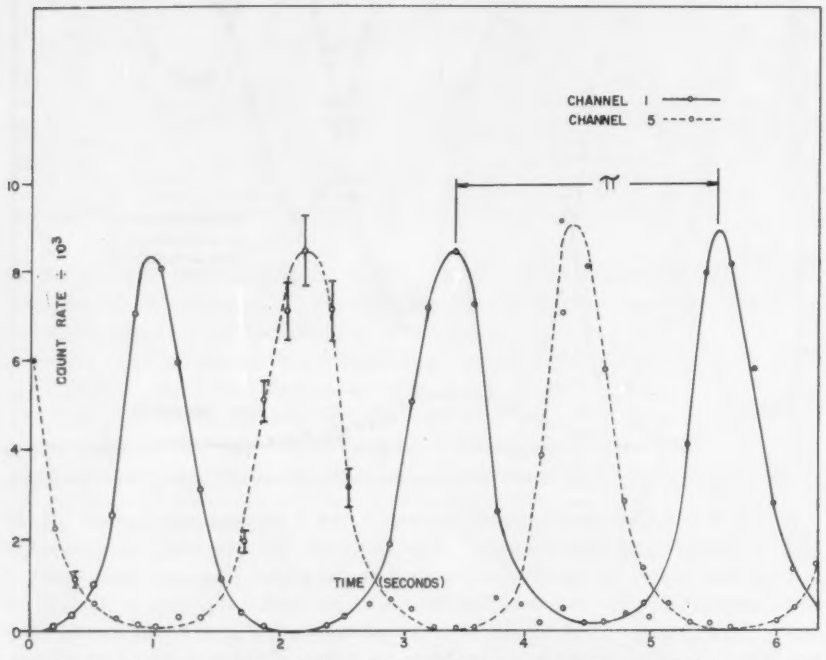

FIG. 8.—Counting rates on channels 1 and 5 for flight 2019. One-half of a revolution of the instrument package is indicated by π. The spin period is 2π.

was very nearly confined to a plane. The maximum half-width at the half height of this angular distribution, consistent with the uncertainties of the measurement, gave an upper limit of 15° for the half-angle of electron distribution about a plane in space. It is shown in the next section that this plane was actually perpendicular to the magnetic field.

The angular distribution near the center of the trapped band was measured with channels 2 and 8, since 1 and 5 were saturated, and 4 had an insufficient counting rate. The altitude of this measurement was 500 km when the package was in the electron band. In Figure 9, counting rates are presented as a function of θ, the

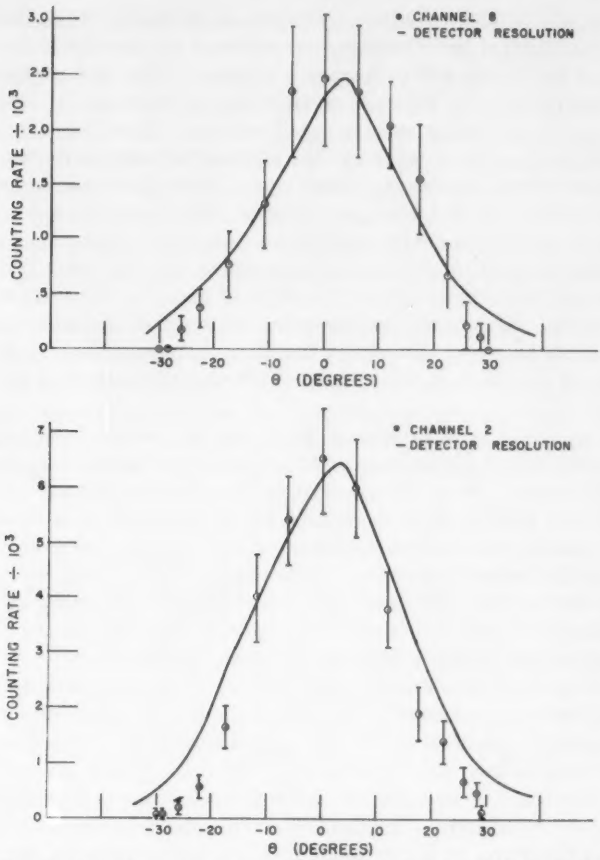

FIG. 9.—Counting rates versus θ for channels 2 and 8 of flight 2019.

angle about the plane of radiation. The smooth curves represent the experimental angular resolutions of the detectors. From these data an upper limit of 15° is derived for the half-angle of the electron distribution. The sharpness of the angular distribution is further illustrated by the fact that the counting rate goes to zero for these counters when they are at an angle of 30° to the plane of the radiation. Thus, for the detector geometries used in this experiment, the electron distribution is indistinguishable from a delta function centered on a plane in space. Since the packages were always spinning and tumbling, it was virtually impossible for the side detectors not to become aligned in the position of maximum counting rate at some part of the cycle. Since the angular resolution of the counters was larger than the observed angular distribution of the flux, only the peak counting rate during a complete spin-tumble cycle was necessary to completely define the field. Phase plots of counting rates (e.g., Fig. 7) for the directional detectors were made for each flight to insure that detectors were oriented in a position to observe the maximum flux.

Orientation of Instrument Package.—Because of the rather strong anisotropy of the radiation pattern of the telemetry transmitter, it was possible to determine the orientation of the instrument package axis in space. The package could be considered a long cylinder for purposes of analyzing its rotation. It had two large principal moments of inertia and one small moment ($^1/_{90}$ of the large moments). Such a configuration has a tendency to transform any angular motion into a flat spin or tumble if there is a means of energy loss. Since the rocket burned out deep in the atmosphere (120,000 feet), aerodynamic forces provided this means. An examination of signal strength records shows a transformation from last stage burnout to a stable mode of tumble established at about 300,000 feet. The flatness of the tumble can be determined from the ratios of moments of inertia and rates of spin and tumble. Since, with the large receiving antennas, it was always possible to receive signals far above the noise, a signal strength record as a function of time may be related to points on the cylindrically symmetric pattern of the telemetry transmitter.

A careful examination of the records allowed one to determine the angle between the longitudinal axis of the package and the line of sight between the package and the receiving station. From this information, it was concluded that the axis of the package lay on a cone about a line of sight having that angle as half angle. Data from other stations provided similar cones. In this manner, the axis of the rocket was found as the intersection of the various cones. Up to five of the six available down range stations could be used at any one time and they gave good agreement for each orientation determination. The stations were located from Patrick to Antigua and covered a look angle from the rocket of almost 150°. As only three stations were necessary to obtain the absolute orientation, the redundancy provided by other stations encourages confidence in the results. This determination was done in selected cases to an accuracy of 3° with respect to a set of axes fixed in space.

To verify that the flux was perpendicular to the lines of force, it was necessary to determine the relation of the axis of the rocket to the field line at the time when counters 1 and 5 were equal and at a relative maximum. Determinations of the direction of the axis of the rocket on favorable flights (3 sec < spin period < tumble period) showed that the plane of the flux and the direction of the line of magnetic field were perpendicular within 3°.

Method of Presentation of Data.—It should now be possible to present that data in the form of peak counting rates versus either altitude or magnetic line. However, it was not possible to choose peak counting rates directly for all portions of a flight because either the counters were too near saturation or the counting rates were so low that a peak counting rate was not statistically reliable. In the cases where a high sensitivity detector was saturated, the counting rate of the low sensitivity detector having the same energy threshold was used. The latter counting rate was normalized in the region where the high sensitivity detector was not yet saturated. In those situations where counting rates of the high sensitivity detectors were too low, the average counting rates of those detectors were plotted and normalized to the peak counting rates where the latter became statistically significant. Corrections for counting losses caused by dead time were applied to all data before normalizations were made.

Two questions may arise concerning these procedures. The first is the validity

of using a low sensitivity detector as a back-up for the high sensitivity detector when the transmissions of the pair were slightly different. This difference produced a negligible error for the energy spectrum observed. The second question concerns the use of average counting rates to extend the peak counting rate curves. For this procedure to be correct, the ratio of average to peak counting rates had to be constant throughout the flight. These ratios were taken from the data at several points along the trajectory of every flight and were found to be constant to within the accuracy of the measurement. This result is to be expected, since the modes of spin and tumble were constant throughout each flight and the direction of the plane of the trapped particles was almost constant over the geographical extent of a rocket flight.

Except for the discussion of absolute flux, the data in this report have been presented as peak counting rates normalized to the high sensitivity detectors: 1 (180 kev threshold), 3 (1 Mev), 6 (4 Mev), and 7 (500 kev). Using the procedures just described, peak counting rates were extended for rates above 2×10^4 counts/sec using the low sensitivity detectors and for rates below about 6×10^2 counts/sec using the average counting rates of the high sensitivity detectors.

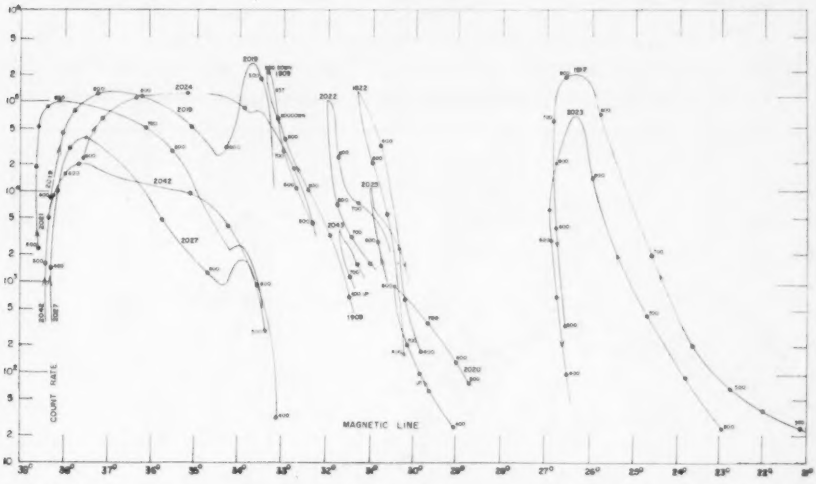

FIG. 10.—Peak count rate versus magnetic line for channel 1 on all successful flights. Hundred kilometer points are indicated.

Plots of peak counting rate against magnetic line for channels 1, 3, and 7 for all flights are shown in Figures 10 through 12. The direction of flight is indicated by arrows. Altitudes of several points for each flight are shown in hundreds of kilometers. In these, as in several succeeding figures, the statistical errors in several selected points are shown by error bars. The Wallops flights are on the left of the figure, the Patrick flights in the middle, and the Ramey flights are on the right. On the flights following event 2, a rather sharp peak is seen to exist at the 33.5° magnetic line; this region is referred to in this report as the band. An expanded view of the band is shown in Figure 13, where again, statistical errors are shown. The band width is shown for Counters 2, 4, and 8; excellent agreement is seen to

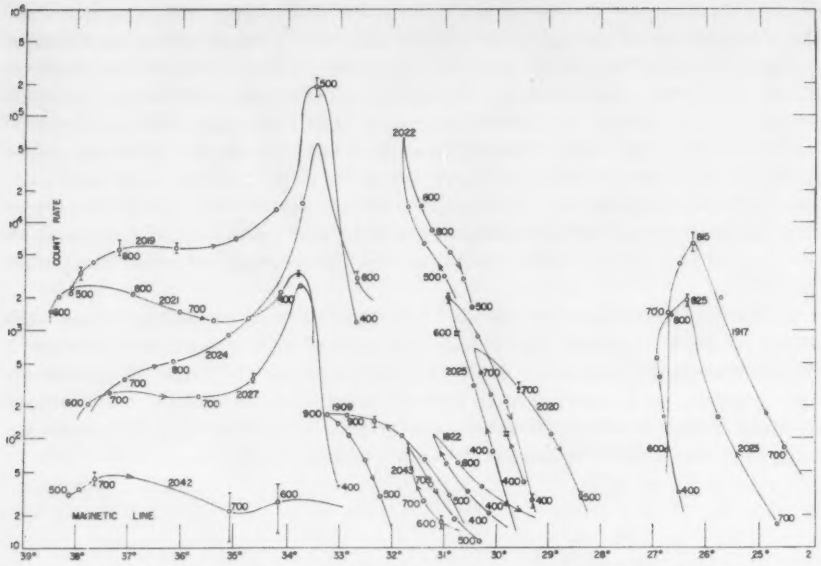

FIG. 11.—Peak count rate versus magnetic line for channel 3 on all successful flights.

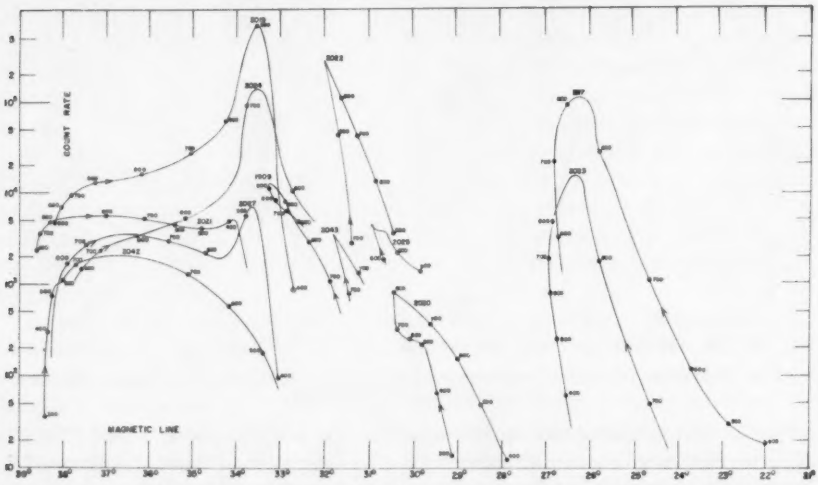

FIG. 12.—Peak count rate versus magnetic line for channel 7 on all successful flights.

exist between these counters. The high count rates observed both to the north and south of the band are known as wings.

Event 1.—Because of the location of the band in event 1, only three sounding rockets were launched. Of these, only two were successful. Even these results are complicated considerably by the fact that 1909, the Patrick flight, went considerably more northerly than any other Patrick flight. Therefore, it is only possible to make

comments on the effects on both sides of the band. Referring to Figures 10 through 12, the Ramey flight (1917) shows a low counting rate until about 600 km on the ascending portion of the flight when the counting rate begins to rise rapidly. The

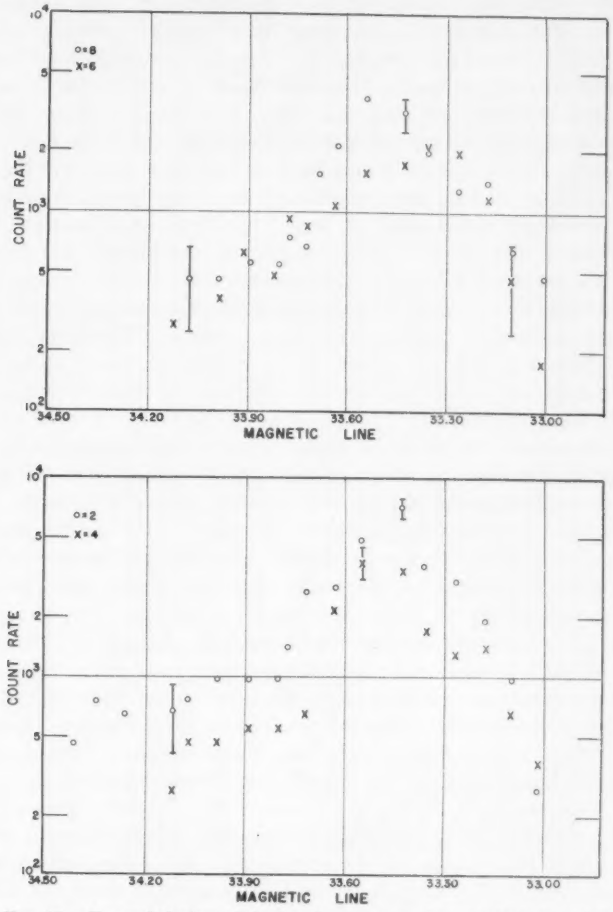

FIG. 13.—Expanded view of the band as seen on channels 2, 4, 6, and 8 of flight 2019.

counting rate continues to increase until apogee after which it recedes only slowly. It may be observed that at the same altitude on the ascending and descending portions of the trajectory, the counting rate is considerably higher at the higher magnetic line. Channel 3 is the most indicative of the presence of any effect added by event 1, as the background for this high energy channel (as illustrated by flight 1822) was much lower than it was in either channels 1 or 7. On Channel 7, the counting rates at similar altitudes are higher on the Ramey flight than on the Patrick flight. However, this effect is even more pronounced on channel 3. Detailed

analysis of effects indicate that the data are consistent with the presence of a band located at the 28° magnetic line plus or minus one degree.

Event 2.—The band from event 2 was more favorably located. The first flight from Wallops after event 2 was launched at H plus 28 minutes. This shows the band clearly and locates it along the 33.5° magnetic line. As a result of this observation, an effort was made to fire north from Patrick; however, because of air traffic problems, it proved to be impossible. Therefore, only the firings from Wallops went directly through the band. However, much valuable information on structure was obtained from the Patrick firings. Because of missile and telemetry failures, there was only one successful firing from Ramey after event 2.

From Figures 10 through 12, it may be seen that the structure of the effect was, in general, a broad plateau upon which was set a narrow band or peak. The location of this band was constant in time. Small apparent changes in location are interpreted as a result of errors in the magnetic field model and the method of projecting the trajectories onto the 75° meridian. Since the magnetic field is not pure dipole at the earth's center, the predominantly eastward electron drift had both small vertical components and north-south components. The simplification of the projection of the trajectory to a purely east-west direction shows up as slight shifts in the plotted position of the band in these figures. The true deviation of the position of the band was less than $^1/_{10}$ of a degree in the period from 40 minutes to 4 days after the event. It should be noted that there was a large area north of the band extending to at least the 38° magnetic line which exhibited counting rates considerably above background though more than an order of magnitude down from the peak and decay properties similar to the band. It is further noted that in this region the counting rate was relatively independent of magnetic line or altitude. A partial explanation of this behavior is that as the package neared the band, the counting rates would tend to increase; however, the package was also descending and the counting rates would normally decrease. These two effects tended to cancel one another, keeping the counting rates fairly constant.

The Patrick flights show a very large effect due to the deviations of our model from the true magnetic field. This is a result of the great longitudinal spread from the eastward fired Patrick flights, and results in some disparity between the ascending and descending portions of the trajectories. However, this does not seriously affect the interpretations of this experiment. The Patrick flights show a large southern wing which extends to at least 29° magnetic line and possibly even further south and which has a shape and decay similar to that of the northern wing. The best measurement of the location of the band was flight 2019, the trajectory of which was almost exactly along the 75° meridian. The data which have been presented were selected from curves smoothed through the raw data.

Spectrum.—The spectrum of the trapped particles was predicted to be a fission spectrum, enhanced in the high energies at late times because of the increased scattering and loss at low energies. This proved not to be the case. Figure 14 shows the spectrum measured in the band for flight 2019 compared with the fission product spectrum. Other flights after the South Atlantic events give the same spectrum within the limits of the data. It is noted that there is a marked deficiency at high energies in the observed spectrum as compared with the fission spectrum.

Lifetime of the Argus Effect.—In order to correlate the results from several

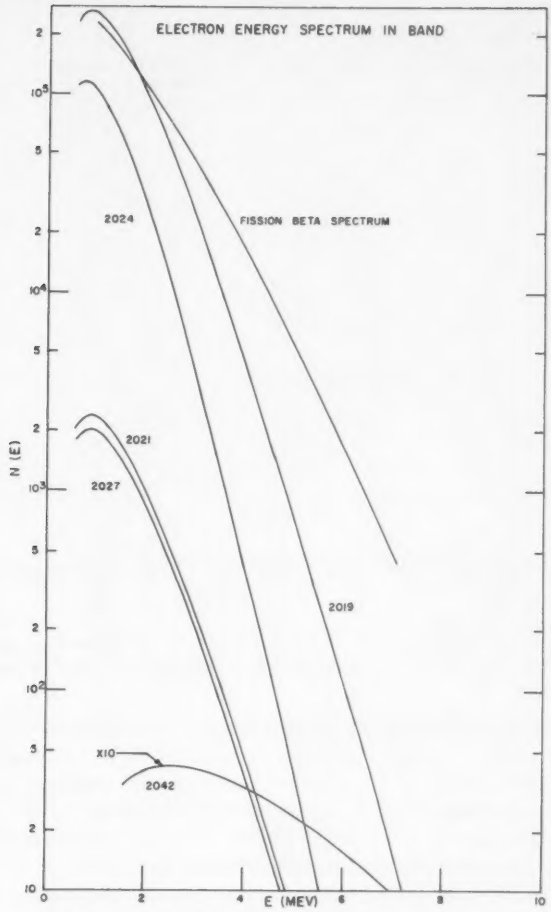

FIG. 14.—Electron energy distribution (electrons per cm² sec Mev) in the bands. The fission beta spectrum is also shown for comparison.

rocket firings, it was necessary to display the data by holding one parameter fixed from flight to flight. The parameter chosen was altitude. A plot of the peak counting rate versus time after event 2 is shown in Figure 15 from the Wallops Island flights. Data are given for counter 3 (1 Mev). The data were selected at a series of constant altitudes taken in the northern wing of the trapped electron shell. From this plot, it can be seen that the effective decay is roughly 1/t. Other plots, i.e., from Patrick firings and on other channels, give basically the same results. On channels 1 and 7, the Wallops flights decay less rapidly since they are also observing a neutron decay beta background from a far earlier time. (See following sections.)

Pacific Events.—The high altitude shots in the Pacific were of large yields and

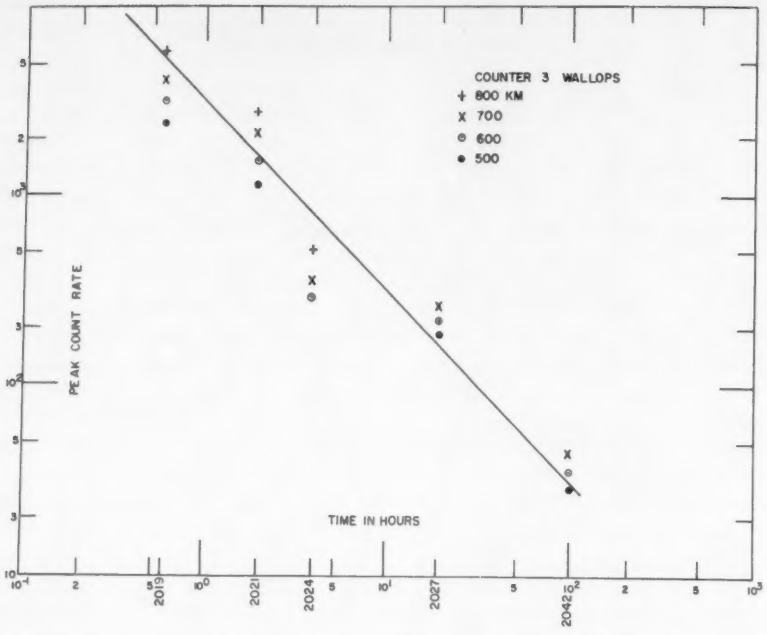

FIG. 15.—Isoaltitude plot of peak count rate on channel 3 versus time after event 2. The data are taken at 100 km intervals on the ascending portion of the Wallops flights.

therefore might be expected to have contributed to the electron backgrounds observed in this project, except that they were at much lower magnetic latitude. Even if it were impossible for these shots to inject fission products into the earth's magnetic field, the large number of neutrons produced would decay, producing betas which might have been trapped. Calculations have been made of the trapped particle density expected from these neutron betas and show that the three missile flights (1822, which took place $2^3/_4$ days after the last Pacific shot; 1909, which, although investigating event 1, was far to the north of the band and may have been affected by the Pacific shots; and 2043, which was long after events 1 and 2 may show the Pacific shots as background. These three flights had very similar characteristics on counter 3, which was not expected to be affected by the neutron decay betas, as its threshold was above the end point of the beta spectrum. Counters 1 and 7, however, should be expected to detect such betas. Flight 1822 was notable for the extremely high counting rates in counter 1. These counting rates are in fact an order of magnitude higher than those at similar altitudes immediately after event 2. Flight 2043 came much later and had considerably lower counting rates on this counter. Channel 7, which failed on flight 1822, was backed up by channel 8, which provided some information. Because of the location of the threshold of these counters with respect to the neutron decay spectrum, it was possible to compare the ratios of the counting rates of these channels with that expected for a neutron beta spectrum. It was found that the ratios of the 180 kev channels to the

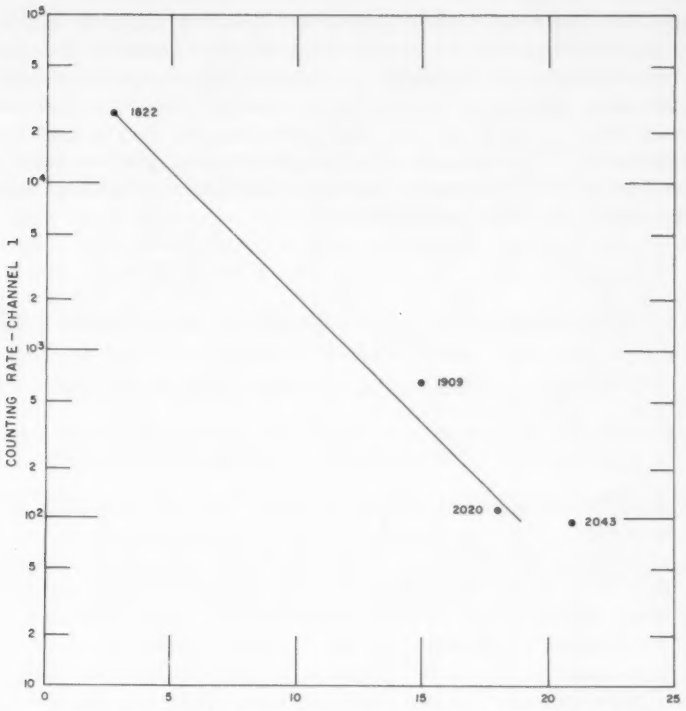

FIG. 16.—Count rate for selected flights as a function of time in days after last Pacific event. All points were taken at 600 km on the ascending portion of the Patrick flights.

460 Kev channels on flight 1822 were those expected for a neutron spectrum within a factor of 1.5.

This is the first experimental verification of the expected effect of the trapping of neutron decay betas from an atmospheric neutron source.

Pacific Lifetimes.—It is difficult to obtain reasonable lifetimes for the neutron decay betas from the Pacific shots, because there was only one flight immediately after a Pacific shot, but before the South Atlantic events. There were many flights, of course, but these were affected by intervening injections. In Figure 16, a plot is shown of peak counting rate on Channel 1 for the Patrick flight versus time after the last Pacific event. It is seen that at long times the decay of the effect is not consistent with the 1/t law which would seem to be valid for short times after event 2. There may be two explanations for this—one, that the particles sometime after 1822, but before 1909, were catastrophically removed by some geophysical event, or that at long times an exponential decay may dominate. Such decay might result from the scattering of electrons out of the shell by hydromagnetic waves.

Conclusions.—The results of the experiment agree qualitatively with those measured by the satellite Explorer IV. If one considers that all losses take place

by small angle scattering in the region about Capetown where the mirror points dip low into the atmosphere as a result of the magnetic anomaly, this experiment agrees quantitatively with the satellite. Certain unique results are obtained from the rocket data: the angular distribution is measured carefully and shown to be distributed about the plane perpendicular to the magnetic lines of force; the location of the band at one point is shown to be quite constant; spectral measurements indicate a deficiency of high energy particles; and the neutron decay betas from the large yield Pacific shots are detected.

THEORY OF GEOMAGNETICALLY TRAPPED ELECTRONS FROM AN ARTIFICIAL SOURCE

BY JASPER A. WELCH, JR., AND WILLIAM A. WHITAKER

PHYSICS DIVISION, RESEARCH DIRECTORATE, AIR FORCE SPECIAL WEAPONS CENTER, AIR RESEARCH AND DEVELOPMENT COMMAND, KIRTLAND AIR FORCE BASE, NEW MEXICO

This paper reports some of the theoretical predictions and interpretations for the Argus experiment. This experiment consisted of three small yield nuclear detonations approximately 300 miles above the South Atlantic Ocean in the late summer of 1958. Beta decay of the fission products from the explosions injected relativistic electrons into trapped orbits in the geomagnetic field. It is the history of these electrons with which we are concerned here. Experimental measurements of this history were made by the satellite Explorer IV and the Jason sounding rockets. Details of these and other measurements performed during the Argus experiment are found in accompanying papers by Van Allen *et al.;* Allen *et al.;* Newman; and Peterson. A general description of the entire experiment is given in the accompanying paper by Christofilos.

Section 1: General Concepts.—We shall take as our point of departure, a source function which is the number of particles per cubic centimeter injected into the geomagnetic field. We shall take the injection to be isotropic. Our first step is to show how the particles rearrange themselves in the field according to the geometry of the trapped orbits. Most of the features are obtained by considering a dipole field and many by even omitting the angular dependence of the dipole. Departures from a static dipole are treated in later sections as perturbations.

We shall make use of the mirror equation of Alfven,[1] $B = B_m \sin^2 \alpha$ where the pitch angle α is the angle between the velocity vector of the particle and the magnetic field vector, where the field strength is B and where B_m is the field strength where the particle is mirrored (i.e., where α becomes $\pi/2$). This is equivalent to the conservation of the magnetic moment of the particle orbit. That is,

$$\mu = (\text{area} \perp \text{to } \vec{B})\ (\text{current} \perp \text{to } \vec{B}) \qquad (1)$$

$$= \pi R_c^2 e \omega_c \sim v_\perp^2 / B$$

where R_c and ω_c are the cyclotron radius and frequency and $v_\perp$ is the component of the velocity perpendicular to B. Alfven has shown that μ is an adiabatic invariant, e.g., not strictly a constant of the motion, so long as space and time

variations of the field are small with respect to the cyclotron radius and period. Section 6 will discuss the consequences of the breakdown of these conditions. Since the static magnetic field does no work on the particle, the total velocity is conserved and we obtain $v_{||}^2 = (v^2 - v_\perp^2) = (1 - B/B_m)v^2$ as the equation of motion of the guiding center along the line of force.

In addition to the particle density $p(r)$, i.e., the number of particles per cubic centimeter, we shall have use for what we call the mirror point density $w(r)$. This is just the number of particles whose guiding centers mirror in a particular cubic centimeter. This concept is quite useful as the lifetime of a particle depends almost exclusively on the location of its mirror point. Or to put it another way, all particles with the same mirror point have the same lifetime (aside from dependence on energy and type of particle); whereas, all particles found at the same altitude, for example, do not have the same lifetime. These three densities—injection, $n(r)$; particle, $p(r)$; and mirror point, $w(r)$—are related in the following ways.

The number of particles injected into a volume of length dl' along the field line at r' having cross sectional area dA', and having pitch angles between α and $\alpha + d\alpha$, is $d\nu' = dA'dl'n(r') \sin \alpha d\alpha$ where α runs from 0 to $\pi/2$ since a particle emitted in the direction $(\pi - \alpha)$ will reappear at the injection point with angle α after one reflection, and vice versa.

These particles will be mirrored in a volume of length $dl = dl/d\alpha d\alpha$ and of cross section $dA = [B(r')/B(r)]\, dA'$ at the point r. But the number of mirror points per unit volume is just the mirror point density. Thus

$$dw(r) = \frac{d\nu'}{dAdl} = \frac{B}{B'} n' \sin \alpha \left|\frac{d\alpha}{dl}\right| dl' \tag{2}$$

where $B = B(r)$, $B' = B(r')$, etc.

But since

$$\sin \alpha \, d\alpha/dl = \frac{d}{dl} \cos \alpha = \frac{d}{dl} \sqrt{1 - B'/B} = \frac{B'}{2B^2\sqrt{1 - B'/B}} \frac{dB}{dl} \tag{3}$$

$$dw = \frac{n'}{2B\sqrt{1 - B'/B}} \frac{dB}{dl} dl' \tag{4}$$

and the total mirror point density is obtained by integrating along the field line of which dl' is the arc line element.

$$w(r) = \frac{1}{2B(r)} \frac{dB(r)}{dl} \int_B^{B^*} \frac{n(r')}{\sqrt{1 - B'/B}} dl \tag{5}$$

The limits refer to the values of $B(r')$; B and B^* are numerically equal to each other and to $B(r)$ and the integral is taken along the field line from the point r in the direction of decreasing field strength and continues past the equator and on until $B(r')$ attains the value $B(r)$ again. This point is called the conjugate point to r; hence, the notation on the limit B^*.

Since we shall find that the mirror point density is most directly affected by losses, let us compute the particle density from mirror point density, although we could go directly from injection density. The time taken by a particle in going

from one mirror point to the conjugate mirror point (let us call this the bounce period) is given by the integral

$$T(v, B_m, I) = \int_{B_m}^{B_m^*} \frac{dl}{v_{||}} = \frac{1}{v} \int_{B_m}^{B_m^*} \frac{dl}{\sqrt{1 - B/B_m}} \tag{6}$$

where B_m is the field at the mirror points and the dependence of T on the particular line of force is denoted by I. The symbol I refers to the integral invariant of the motion $I = \int_{B_m}^{B_m^*} v_{||} dl$ and is more fully developed in Section 3. For a dipole the equation of a field line is $r = r_0 \sin^2 \theta$ where θ is the co-latitude, r is the distance from the dipole (i.e., we use the spherical polar coordinates r, θ, ϕ), and r_0 is the value of r at the equator. We shall call r_0 the field line constant.

The bounce period for a dipole field is very nearly proportional to the field line constant r_0, and independent of the angle through which the particle travels. The numerical values for relativistic electrons for an r_0 of two radii is about 0.2 second.

Let $d\nu = w' dA' dl'$ be the number of particles with mirror points in the volume of length dl' and cross section dA' at r'. These particles will spend the fraction of their bounce period dl $(v_{||} T')$ in the volume element of length dl and cross section $dA = dA'$ (B'/B) at the point r. Thus we have

$$dp = \frac{d\nu (dl/v_{||} T')}{dl dA} = \frac{w' B}{v_{||} T' B'} dl' \tag{7}$$

$$p(r) = \frac{B(r)}{v} \int_B^{B_{max}} \frac{w' dl'}{B' T' \sqrt{1 - B/B'}} \tag{8}$$

where the integral is taken along the field line in the direction of increasing B from the point r and continuing until w' goes to zero for good (say at the earth's surface). As we noted before, although T' is strictly a variable, it is slowly varying for any particular field line and hence can usually be taken out of the integral.

We would like to point out the radical in the denominator of the integrand for both p and w goes to zero at the limit of integration. While not making the integral blow-up, the radical does tend to cause most of the contribution to the mirror point density w to come from injection points nearby and likewise, most of the contribution to the particle density p to come from mirror points nearby.

Several other useful integrals—angular distribution of the particle flux, flux parallel to B, etc.—will be developed by similar methods in Section 5.

For our source function, we shall take electrons injected with constant injection density into a tube of force. We shall consider a dipole field and let the tube lie along the line of force that intersects the earth at 45° magnetic latitude. The tube will have a circular cross section. The cross section will be inversely proportional to total field strength. The mirror point density is then calculated according to equation (5),

$$w(r) = \frac{n_0}{2B} \frac{dB}{dl} \int_B^{B^*} \frac{dl'}{\sqrt{1 - B'/B}}$$

which is plotted in Figure 1.

We will later need the arc length from the equator along the field line $r = r_0 \cos^2 \lambda$ where λ is now the latitude. This is given for a dipole field by

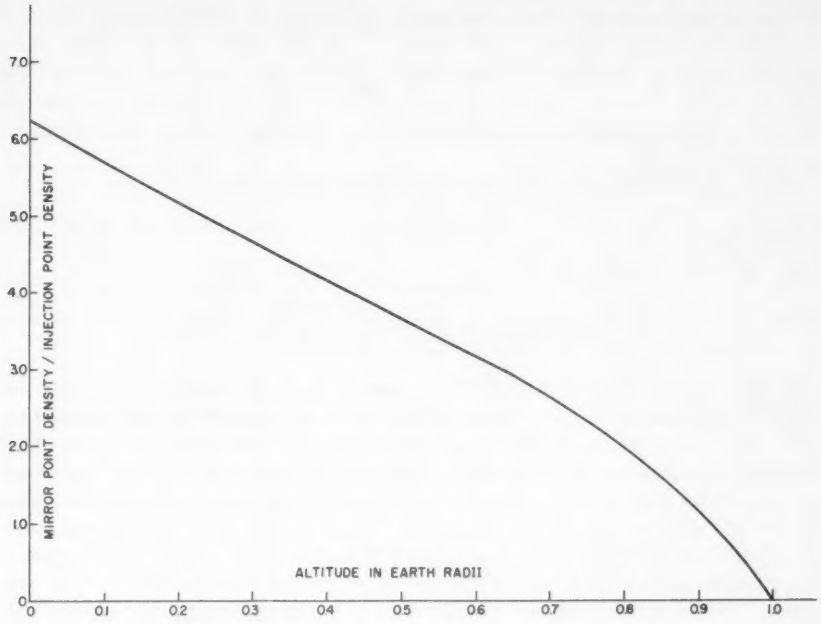

FIG. 1.—Initial mirror point density versus altitude.

$$s(\lambda) = \frac{r_0}{2\sqrt{3}} \left\{ (3\sin^2\lambda)^{1/2} (1 + 3\sin^2\lambda)^{1/2} + ln\left| (3\sin^2\gamma)^{1/2} - (1 + 3\sin^2\lambda)^{1/2} \right| \right\}$$

which can be approximated for $\lambda \leq \pi/4$ by

$$S(\lambda) = r_0 (\lambda + \lambda^2/10)$$

Section 2: The Drift in Longitude.—This tube filled with electrons does not however, remain stationary. There are several phenomena which give rise to drifts. First, there is the gradient of the magnetic field. As given in Spitzer,[2] this drift velocity is in the direction $\nabla\vec{B} \times \vec{B}$ for electrons (i.e., east) and has the value

$$V_{DG} = v_\perp \frac{R_c \nabla_\perp B}{2B} = \frac{v_\perp R_c}{2} \frac{1}{\Re} = v_\perp{}^2 \frac{1}{2\Re w_c}$$

where

V_{DG} = drift velocity due to gradient of magnetic field
$v_\perp$ = particle velocity perpendicular to B, and
$\nabla_\perp B$ = gradient of the scalar B in plane perpendicular to $\vec{B}$.

We have replaced $B/\nabla_\perp B$ by $\Re$, the radius of curvature of the field line, which is valid if curl $\vec{B} = 0$.

Second, there is a drift if the particle is moving with a velocity $v_{//}$ along a line

of force which is curved. From Spitzer this gives rise to a drift velocity,

$$V_{DC} = v_{\parallel}^2 \frac{1}{\Re \omega_c},$$

where $v_{\parallel}$ = particle velocity parallel to field line. This drift is in the same direction as the gradient drift.

Third, there is the drift from the earth's gravitational field,

$$V_{DE} = \mathfrak{g}/\omega_c.$$

This is very small as

$$\frac{V_{DE}}{V_{DC}} = \frac{\mathfrak{g}/\omega_c}{v_{\parallel}^2/\Re\omega_c} = \frac{\mathfrak{g}\Re}{v_{\parallel}^2} \backsimeq \frac{10^3 \times 10^9}{(3 \times 10^{10})^2} \backsimeq 10^{-9}.$$

Fourth, the direction of the drifts from the gradient and curvature are west for positively charged particles. Thus, the electrons and positive ions would tend to pull apart until the resulting charge separation built up enough electric field strength to overcome the drift forces. This electric field will be in the east-west or ϕ direction and will itself cause a drift radially outward. If unchecked, this drift would tend to disperse the electrons. However, the earth is supposed to be immersed in a hydrogen plasma orginating from the solor corona. This plasma has an estimated 10^3 protons and thermal electrons per cubic centimeter. If this is true, the electric field will never build up as this comparatively dense plasma will shift around to neutralize the charge separation. The uncertainties about this point and the presence and effect of electric fields in general seem to be satisfactorily dispelled by the observed stability of the Argus electron shell.

The curvature for a dipole field is given by

$$\frac{1}{\Re} = \frac{3}{r_0 \sin\theta}\left[\frac{1 + \cos^2\theta}{(1 + 3\cos^2\theta)^{3/2}}\right]$$

and the cyclotron frequency by

$$\frac{1}{\omega_c} = \left(\frac{\gamma m_0 c}{eB_0 r_e^3}\right)\left(\frac{r_0^3 \sin^6\theta}{(1 + 3\cos^2\theta)^{1/2}}\right)$$

and the velocity term by

$$v_{\parallel}^2 + \frac{1}{2} v_{\perp}^2 = v^2 - \frac{1}{2} v_{\perp}^2 = \beta^2 c^2\left[1 - \frac{B}{2B_m}\right].$$

But we really want the drift velocity in radians per second, $\dot{\phi} = V_D/r \sin\theta = V_D/r_0 \sin^3\theta$. Combining we obtain

$$\dot{\phi} = \left(\frac{r_0\beta^2\gamma}{r_e}\right)\left(\frac{3m_0c^2}{eB_0r_e}\right)\left(\frac{c}{r_e}\right)\left\{\left[\frac{1 + \cos^2\theta}{(1 + 3\cos^2\theta)^2}\right][1 - B/2B_m]\ [\sin^2\theta]\right\}.$$

The numerical constant is easily evaluated as

$$\left(\frac{3m_0c^2}{eB_0r_e}\right)\left(\frac{c}{r_e}\right) = \left(\frac{1.53\ \text{Mev}}{6.02 \times 10^4\ \text{Mev}}\right)\left(\frac{3 \times 10^{10}\ \text{cm/sec}}{6.38 \times 10^8\ \text{cm}}\right) = 1.20 \times 10^{-3}\ \text{sec}^{-1}.$$

We must remember that B/B_m is also a function of θ and θ_m, the colatitude of the mirror point. For our case the line intersects the earth at $\theta = 45°$, $r_0 = 2r_e$ and we may easily evaluate $\dot{\phi}$ for the case of particles mirroring on the equator. In this case we have $\theta_m = \pi/2$ and $B = B_m$ always. Then

$$\dot{\phi} = 1.20 \times 10^{-3}(2\beta^2\gamma)\{[1]\ [1/2]\ [1]\}$$
$$= 1.20 \times 10^{-3}\beta^2\gamma \text{ sec}^{-1}.$$

The drift period is then

$$T_D(\theta_m) = T_D\left(\frac{\pi}{2}\right) = 2\pi/\dot{\phi} = 5.23 \times 10^3/\beta^2\gamma \text{ sec}$$
$$= 87/\beta^2\gamma \text{ minutes}$$
$$\approx 30/E_{mev} \text{ minutes}$$

For other values of θ_m we need the average $\dot{\phi}$ weighted by $v/v_{||}$, the relative time spent at each angle θ. This weighted average value turns out to vary by less than 8 per cent over the range $\pi/6 \leq \theta_m \leq \pi/2$. Using an average value over θ_m we have constructed Table 1.

TABLE 1

TIME TO DRIFT AROUND THE WORLD VERSUS ENERGY

E, Mev	T average, minutes/revolution
0.5	57
1.0	33
2.0	18
3.0	12.5
4.0	9.7
6.0	6.6
8.0	5.2

The energy dependence of the drift has the effect of a gigantic time of flight spectrometer. If we station ourselves at the eastern coast of the United States (which is almost one revolution from the initial tube for the eastward drifting electrons) and observe at such a time that the 1 Mev electrons just get to our station, we will also see 2 Mev electrons that have gone around twice and 3 Mev electrons that have gone around three times, and so fourth.

When we add up all the contributions we will get a total flux above 1 Mev that varies in time as shown in Figure 2. This figure is for a fission product beta spectrum which drops off rapidly above 1 Mev. The large discontinuities occur just as the large 1 Mev contribution passes by the station. For any detector one will have some time variation of the flux associated with the energy dependent drift and it will take about six drift periods before the spectrum is completely smeared out. Thus, at early times some care needs to be taken in the proper interpretation of the flux levels.

Section 3: The Shape of the Electron Shell.—For a pure dipole field, or any other field where B has no components in the ϕ (i.e., azimuthal) direction, the drift will be entirely in the ϕ direction. However, the geomagnetic field does have ϕ components even when referred to an eccentric dipole. These components were of some concern to us in locating our rocket launch sites and trajectories. We

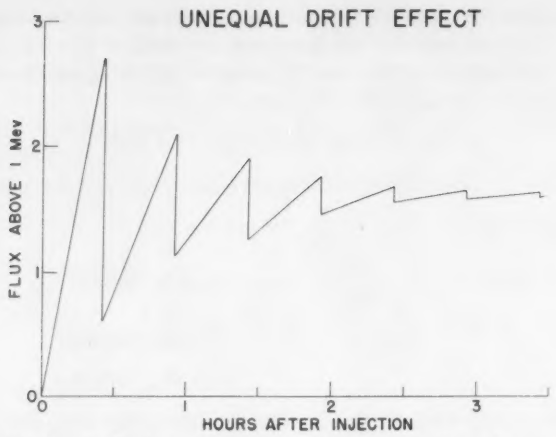

FIG. 2.—Total electron flux above 1 Mev versus time after an injection.

needed an estimate of what we call the conjugate trace—the intersection of the electron shell with the earth's surface. We shall define magnetic latitude as latitude with respect to the position and direction of the best fit (eccentric) dipole.

If the field is pure dipole field, then the conjugate point trace will follow a constant magnetic latitude with respect to the dipole. Since the eccentric dipole is located a small fraction of the earth's radius from the earth's center, we can lay off constant latitude approximately as small circles from the point of emergence of the eccentric dipole axis from the earth's surface.

If we examine a plot of the field at the earth's surface, we see that there are moderate deviations from the dipole approximation. We shall now give two estimates of the effect of these perturbations on the trace of the conjugate point.

As we go away from the earth's surface, the surface perturbations can be expected to wash out and the field to approach that of a dipole. Calculation of the eastward drift for a dipole field shows that most of the drift occurs at the maximum altitude. Further, for a dipole field, the drift has no radial components. Carrying these tendencies to the limit, we can form a model where the electrons always orbit about lines which cross the equatorial plane at a fixed distance from the dipole location. We need now to find where these lines intersect the earth. Now, if the field were pure dipole, these lines would have a constant dip angle at the earth's surface and would intersect the earth along a parallel of magnetic latitude. That is, the conjugate trace, the magnetic parallel, and the iso-dip line would be coincident.

For a dipole field, the dip angle D (the angle between the line of force and the earth's surface) is related to the latitude, λ, by $\tan D = 2 \tan \lambda$. Thus, the dip angle associated with 45° latitude is 63°. Consider now a particular longitude in the northern hemisphere where high-order multipoles have caused the 63° iso-dip trace to lie south of 45° magnetic latitude (see Fig. 3). From the figure we see that the line of force intersecting the earth at 45° latitude for a pure dipole field now intersects at a lower latitude, but not so low as the 63° iso-dip trace. This new intersection is just the position of the conjugate trace at this longitude.

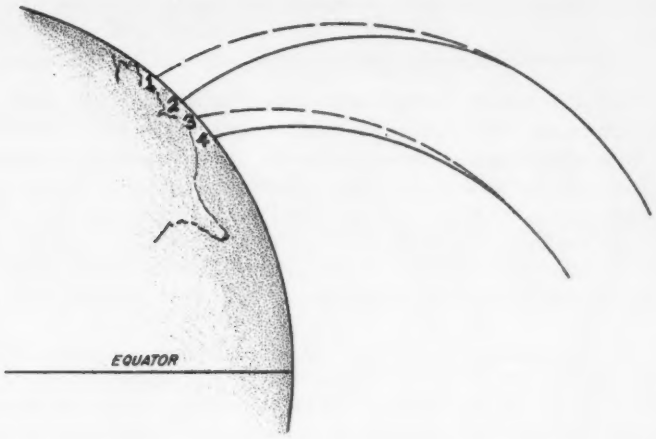

FIG. 3.—Illustration of the earth's magnetic field along the U.S. east coast. Point 1 is 44-degree magnetic latitude and unperturbed position of 60-degree iso-dip line. Point 2 is perturbed position of 1, i.e., the conjugate trace. Point 3 is the unperturbed position of line that is now at Point 4. Point 4 is 60-degree iso-dip line, perturbed position. Note that the conjugate trace lies between 44 degrees magnetic latitude and the 60-degree iso-dip line. For a dipole field all three would be coincident.

The first estimate is to place the conjugate trace halfway between a parallel of magnetic latitude and the perturbed iso-dip trace characteristic of this latitude for a dipole field (e.g., 63° iso-dip trace for 45° latitude). There are generally a large number of latitudes and corresponding iso-dip traces which lie equally far north and south of a given initial conjugate point. The pair with smallest separation will probably give the best results. The conjugate trace as seen by the Explorer IV agreed quite well with this approximation.

A second approximation is to assume that all the drift takes place at or near the earth's surface. Then the direction of the drift is always perpendicular to both the magnetic field and that part of the gradient of the magnetic field perpendicular to the field itself (roughly magnetic east). Thus, we can take the published values for the magnetic field and find this direction for all points on the earth. Starting from the initial conjugate point we can then wind our way around the earth, guided by the direction of the drift as described above. This approximation gives results quite close to those from the iso-dip approximation.

A more elegant method is to make use of the two adiabatic invariants of the particle motion—the magnetic moment and the integral invariant. The constancy of magnetic moment implies that a given particle will always mirror at a constant value of the field strength. Thus, insofar as loss times are long compared to drift times, we should find the same number of electrons in the shell at various

longitudes if we keep on a surface of constant field strength. The integral invariant, $I = \int_{Bm}^{B_m^*} v_{||} dl$, is conserved since the bounce period (<1 sec) is small compared to the azimuthal drift period (~1000 sec). Thus, to construct the shell at each longitude, one moves north and south along a surface of constant field strength until the integral invariant comes out correctly. For azimuthally distorted fields particles with different mirror point field strengths that share a common field line at one longitude may not share a common field line at other longitudes. That is, the shell need not have a constant thickness as a function of longitude, even on a surface constant field strength. Wherever the shell is wider, however, the electron density will be lower, and vice versa. This effect is noted in the satellite data.

Section 4: Atmospheric Losses and the Characteristic Lifetime.—Small-angle scattering: It has been shown by Christofilos[3] that small-angle scattering of electrons by the air is a more important loss mechanism than energy loss by either radiation or ionization. We shall see in this section that the role of scattering is to move the mirror point of the electrons to lower and lower altitudes. As they encounter more dense air, the electrons eventually lose all their energy by ionization.

On the basis of small-angle scattering, the mean lifetime of an electron as a function of its initial mirror point altitude will now be calculated. We will call this the characteristic lifetime. Many of the approximations will be based on the case of those lines of force that intersect the earth's surface near 45° latitude. Here the dip angle is large, ≈70°, and motion parallel to the lines of force (near the mirror point) may be approximated to motion along an earth radius. At this latitude the angular factor in the field strength, $(1 + 3 \cos^2 \theta)^{1/2}$, has increased to only 1.5 from its value of unity at the equator. Because of the slow change and the large dip angle we will refer all field strengths to their value at the mirror point and keep only the $1/r^3$ dependence in field strength. To account for the effect of the eccentric dipole we shall construct a model atmosphere in which the air density has been averaged over longitude. These densities will be referred to the altitude of closest approach to the earth's surface. For 45° magnetic latitude and 100-*km* scale height the averaged density is down by a factor of 3.6 from the density at closest approach.

Using this approximation to the field the mirror equation becomes $r_m^3 = r^3 \sin^2 \alpha$, where the subscript m denotes a value at the mirror point. Now physically, scattering of the electron by the atmosphere represents a change in α at constant r; hence, a change in r_m. Let ϵ be an elemental scattering angle. Expanding $\sin^2 (\alpha + \epsilon)$ in a Taylor series about α

$$\sin^2 (\alpha + \epsilon) = \sin^2 \alpha + \sin(2\alpha)\epsilon + \cos(2\alpha)\epsilon^2 + \ldots .$$

But we are interested in the average behavior of the particles and ϵ is distributed in a symmetric, random distribution about $\epsilon = 0$. Thus, let us average over this distribution. Then we have

$$\langle \sin^2(\alpha + \epsilon) \rangle = \sin^2\alpha + \cos (2\alpha) \langle \epsilon^2 \rangle$$

neglecting higher terms. Let the symbol D denote a differential averaged over the

distribution in ϵ. Then we have

$$D\,(\sin^2\alpha) = \cos\,(2\alpha)\,\langle\epsilon^2\rangle.$$

But from Fermi[4] we obtain

$$\langle\epsilon^2\rangle = \frac{7 \times 10^{-3}\,D(s)}{E^2}$$

where $D(s)$ is the elemental path length in STP air, and E in the electron energy in Mev. For relativistic electrons,

$$\langle\epsilon^2\rangle = \frac{7 \times 10^{-3} NcD(t)}{E^2 N_{\mathrm{STP}}} = KND(t)$$

where N is the atomic particle density where the scattering occurs (we will be interested in altitudes for several hundred miles, hence the choice of atomic particle density), c is the velocity of the electrons set equal to light velocity. Remembering that D refers to an event happening at constant r,

$$D(r_{\mathrm{m}}{}^3) = 3r^2{}_{\mathrm{m}} D(r_{\mathrm{m}}) = r^3 D(\sin^2\alpha)$$

or

$$D(r_{\mathrm{m}}) = \frac{r^3}{3r^2{}_{\mathrm{m}}} \cos\,(2\alpha)\,KND(t).$$

We can then define a velocity of the mirror point of the mean scattered particle

$$V' \equiv \frac{D(r_{\mathrm{m}})}{D(t)} = \frac{r^3}{3r^2{}_{\mathrm{m}}} \cos\,(2\alpha)\,KN.$$

Let us now calculate what is of more interest, the average of V' over the path of the particle. If the scattering per trip is small we can just take the average over the unperturbed path

$$V = \frac{\int_0^{\mathrm{T}} V'dt}{\int_0^{\mathrm{T}} dt}$$

where the integrals are taken over the path for the mirror point of the equator.

$$Y \equiv \int_0^{\mathrm{T}} V'dt = \int_{r_{\mathrm{m}}}^{r_0} \left(\frac{r^3}{3r^2{}_{\mathrm{m}}}\right) \cos\,(2\alpha) KN(r)\,dt.$$

Now $N(r) = N_{\mathrm{m}} \exp\,(-x/a)$ where $x = (r - r_{\mathrm{m}})$ and a is the e-folding distance for the atmospheric density. We can express all variables inside the integral in terms of x, expanded about r_{m}.

$$\cos\,(2\alpha) = -1 + 6\,(x/r_{\mathrm{m}}) + \ldots$$

$$r^3/3r^2{}_{\mathrm{m}} = (r_{\mathrm{m}}/3)(1 + (x/r_{\mathrm{m}}) + \ldots)$$

$$\frac{dv_{\parallel}}{dt} \equiv u(r, r_{\mathrm{m}}) = u(x, r_{\mathrm{m}}) = (3c^2/2r_{\mathrm{m}})(1 - 4(x/r_{\mathrm{m}})\ldots)$$

where u is the acceleration along the line of force.

Since at the altitudes of most interest, a few hundred miles, $a \ll r_m$ and the exponential factor is dominant. Specializing to the case where $(r_0 - r_m)$ is also much larger than a, we can express everything in terms of x and make the limits of integration zero to infinity.

Then Y becomes

$$Y = -\frac{r_m K N_m}{3(2u_m)^{1/2}} \int_0^\infty \frac{\exp(-x/a)}{\sqrt{x}}\, ds = -\frac{r_m K N_m}{3(2u_m)^{1/2}} \sqrt{a\pi},\ u_m = \frac{3c^2}{2r_m}$$

expressing

$$T = f(2r_m/u_m)^{1/2} = f\sqrt{4/3}(r_m/c)$$

we find that $f = 2.0$ almost independent of r_m. Thus

$$V = Y/T = -\frac{\sqrt{\pi} K N_m}{6f} (ar_m)^{1/2}.$$

Figure 4 shows how the air density throughout a bounce period is appreciable only when $\cos 2\alpha$ is close to -1. That is, since all scatters occur when α is nearly $\pi/2$ all scatters lead to a lowering of the mirror point. The statistical nature of the scattering relates then only to the amount of lowering.

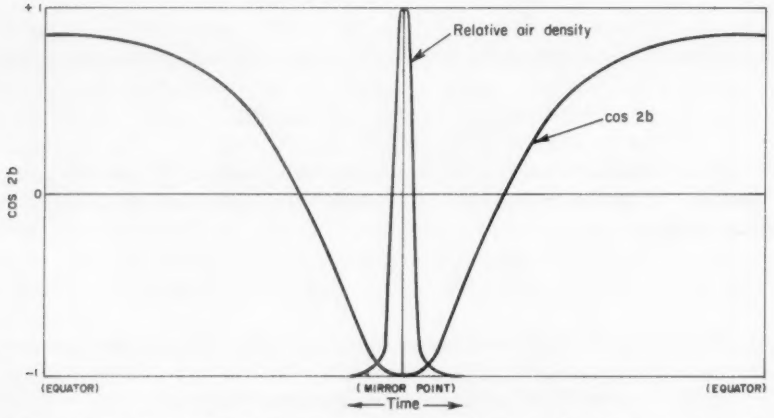

FIG. 4.—Relative air density encountered by a particle mirrored at 45 degrees magnetic latitude versus time. Also $\cos 2\alpha$ where α is the angle between the velocity vector and the magnetic field.

It can be seen that the average velocity is proportional to the density at the mirror point and is downward into the exponentially increasing atmosphere. Under these circumstances, and again neglecting the slow variation of a and r_m compared to N, we find the characteristic lifetime to be completely lost is the e-folding distance, a, divided by the initial downward velocity. To show this, let $V(x)$ be the velocity positively downward as a function of $x = (r_i - r_m)$, where r_i is the initial mirror point. Then $V(x) = V(O) \exp(x/a)$ neglecting the variation

of V with $(ar_m)^{1/2}$. Then the characteristic lifetime is

$$\tau(r) \cong \int_0^\infty \frac{dx}{V(x)} = \frac{1}{V(0)} \int_0^\infty \exp(-x/a)\, dx = \frac{a}{V(0)}.$$

Substituting for K and setting $18/\sqrt{\pi} = 10$,

$$\tau(r) = \frac{10^7 E^2 f}{N} \left(\frac{a}{r}\right)^{1/2} \doteq \frac{2 \times 10^6 E^2}{N} \text{ days.}$$

This result is valid for $(r_0 - r_t) \gg a$ and, $(r_0 - r_t) \approx r_t$. This equation has been evaluated in Table 2 for the Jastrow and Harris Atmosphere[5] (curve A).

Time variation of the mirror point density: We have calculated in the previous section, the mean time for a particle's mirror point to drift down to the earth. This process was controlled by small angle scattering. In this section we shall investigate the cumulative effect on the mirror point density at a given altitude from the mirror points that are lost to lower altitudes and those that are gained from higher altitudes. We shall define $\tilde{w}$ as the integral of the mirror point density w over the cross sectional area of the shell perpendicular to the field line.

We have seen that the mean velocity of a mirror point downward is proportional to the air density, $N(h)$, at the mirror altitude, h. That is, $dh/dt = -K_1 N(h)$ where K_1 is a constant. Let us approximate the function $N(h)$ by an exponential, $N(h) = N_0 e^{-h/a}$ where N_0 and a are constants chosen to fit the density in the region of interest. Now define a new variable, z, such that

$$dz = dh e^{+h/a}; \qquad z = \int_0^h dh e^{h/a} = a[e^{h/a} - 1]$$

$$h = a \ln\left(\frac{z + a}{a}\right); \quad dh/dz = a/(z + a).$$

Then we obtain $dz/dt = -K_1 N_0 = -V_0$, where V_0 is a constant. That is, in "z - space" the mean velocity downward is constant. Let us now make the approximation that all mirror points drift with the mean drift velocity. This assumption will have the effect of maintaining peaks and valleys in the mirror point density. Physically, they should wash out because of the statistical nature of the scattering process. However, we have seen in Section 1 that we do not expect peaks and valleys for the initial condition of our particular problem. Paul Nakada at Livermore has checked this in some detail and has applied more rigorous scattering theories. He obtains essentially the same results.

Proceeding then, our assumption implies that $\tilde{w}(z, t) = \tilde{w}(z - V_0 t, 0)$ where $w(z, t)$, is the mirror point density in "z-space" at time t, and is related to the density in "h-space" by $\tilde{w}(z, t)dz = \tilde{w}(h, t)dh$. We can now find the time variation of $\tilde{w}(h, t) = \tilde{w}(z, t)(dz/dh)_z = \tilde{w}(z + V_0 t, 0)(dz/dh)_z$ as $\tilde{w}(h, t) = \tilde{w}(h', 0)(dh/dz)_{z - V_0 t}$ $(dz/dh)_z$. where $h' = h(z + V_0 t)$

$$\tilde{w}(h, t) = \tilde{w}(h', 0) \left(\frac{a}{z + V_0 t + a}\right)\left(\frac{z + a}{a}\right) = \tilde{w}(h', 0) \left(\frac{1}{1 + \dfrac{+ V_0}{a + z} t}\right)$$

but

$$+\frac{V_0}{a+z}=\frac{V_0}{a+ae^{h/a}-a}=\frac{V_0e^{-h/a}}{a}\equiv\frac{1}{\tau(h)}.$$

Therefore we have the result

$$\tilde{w}(h,\,t)\,=\,\tilde{w}(h',\,0)\left[\frac{1}{1+t/\tau(h)}\right]$$

where

$$h'\,=\,h+a\ln\left(1+\frac{t}{\tau(h)}\right).$$

Time variation of the electron spectrum: For the initial beta-spectrum we make use of the neutrino research which has developed semi-theoretical and experimental methods to find the total beta-spectrum for reactors. Muehlhause and Oleska[6] give a spectrum which is very nearly $P(E)\,=\,P_0\exp(-E/1.3)$ for $0.5\text{ Mev}<\text{E}<6$ Mev. We found $\tau\,\approx\,E^2$ and

$$\tilde{w}(h,\,t)\,=\,\tilde{w}(h',\,0)\left[\frac{1}{1+t/\tau}\right].$$

Using these and setting $w(h',\,0)\,=\,1$ we can compute the spectrum as a function of time. This has been done in Figure 5, where times are measured in units of the characteristic lifetime at 1 Mev.

We can integrate these spectra to get the decrease in the total mirror point density. This is plotted in Figure 6, where the time units are again the characteristic time for 1 Mev. At long times, this goes as

$$\tilde{w}(h,\,t)\,\rightarrow\,\tilde{w}(h',\,0)\,\frac{5.2\tau(E\,=\,1)}{t}\text{ i.e., for }t\gg\tau(E\,=\,1)$$

which shows that the mean energy of the long-lived electrons is

$$\sqrt{5.2}\,=\,2.3\text{ Mev}.$$

This "$1/t$ law" has been verified fairly well by the rocket and satellite results. One difficulty, however, is that no evidence of turning over at early times (i.e., $t<\tau(h)$) was observed. Some explanation of this may be tied into the effect of the South Atlantic anomaly as discussed in Section 7.

For times greater than the characteristic time, τ (itself a function of altitude and energy), the mirror point density is just inversely proportional to average air density by continuity arguments. This will simplify some of the integrals in the next section.

Section 5: Electron Flux and Angular Distributions.—The isotropic flux is of course just the product of particle density and mean velocity. This product can be obtained from the mirror point density by the formula developed in Section 1:

$$\rho v\,=\,B\int_B^{B_{max}}\frac{w'\,dl'}{B'T'\sqrt{1-B/B'}}.$$

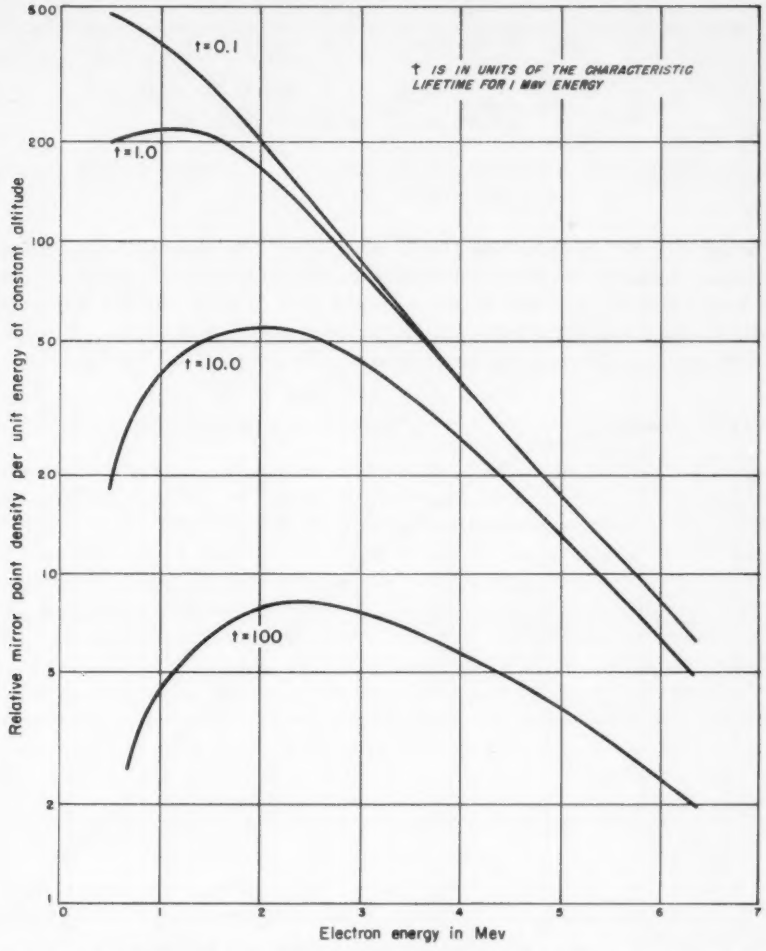

FIG. 5.—Change in the electron spectrum with time caused by the dependence of the characteristic lifetime on electron energy.

Again expanding in terms of $x = r - r_m$, setting $T' = T$, a constant, and setting $w = w_0 e^{-x/a}$ we get $\rho v = \sqrt{\pi a r/3}\,(w_0/T)$. This shows the almost direct relationship of flux on mirror point density.

The angular distribution of the flux is obtained by starting with the differential flux.

$$v d\rho = \frac{v w' B}{v T' B'}\, dl' = \frac{v w' B}{v_{\parallel} T' B'} \frac{dl'}{d\alpha}\, d\alpha.$$

The flux per steradian at angle α is then

$$\frac{dj}{d\Omega} = \frac{vd\rho}{2\pi \sin \alpha d\alpha} = \frac{vw'B}{v_{\parallel}T'B'} \frac{dl'}{d\alpha} \frac{1}{2\pi \sin \alpha}$$

$$= \frac{vw'B}{\pi T'B'v_{\parallel}\sqrt{1 - B/B'}} \frac{(B')^2\sqrt{1 - B/B'}}{BdB'/dl'}$$

$$= \frac{vw'B'}{\pi T'v_{\parallel}} \frac{dl'}{dB'}.$$

Although it is not evident from this formula, if w' is exponential with a 100 km relaxation distance, the angular distribution will be symmetric about $\alpha = \pi/2$ and have a half width at half height of around 11°. That is, the flux will be very nearly confined to a plane perpendicular to the magnetic field vector. Even if w' is fairly constant, the $v_{\parallel}$ in the denominator tends to emphasize the flux near $\alpha = \pi/2$.

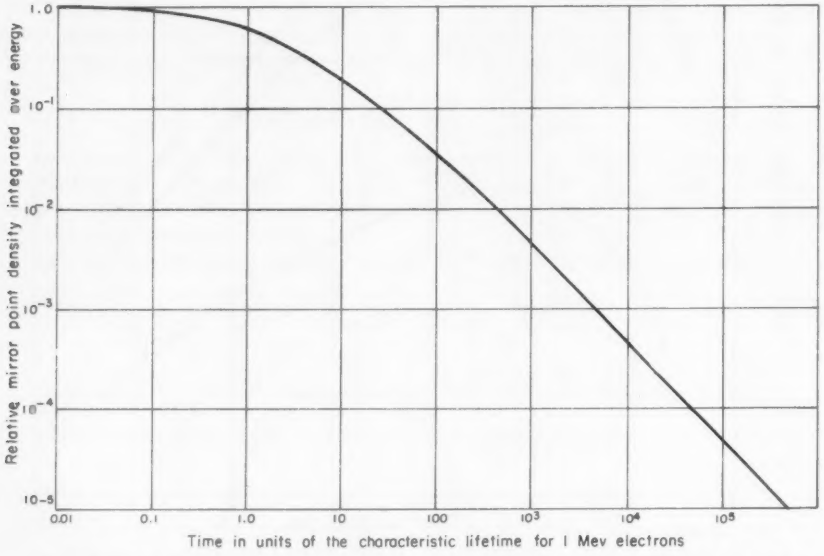

FIG. 6.—Change in the mirror point density integrated over energy with time at constant altitude.

Another interesting integral is the flux parallel to the magnetic field. This is

$$j_{\parallel} = \rho v_{\parallel} = \frac{B}{v} \int \frac{v\sqrt{1 - B/B'}}{B'T} \frac{w'dl'}{\sqrt{1 - B/B'}} = \frac{B}{T} \int \frac{w'dl'}{B'} \simeq \frac{1}{T} \int w'dl'$$

if w' is a rapidly decreasing function of B'. That is, $j_{\parallel}TA$, where A is the cross sectional area of the entire shell, is just $A \int w'dl'$, the total number of mirror points below the point of observation. Now $j_{\parallel}$ is not an easy quantity to measure, but it is related to $j_{\perp}$ through the width of the angular distribution.

Section 6: Magnetic Scattering.—Let us investigate how much space and time

variations in the geomagnetic field might affect the conservation of the magnetic moment. That is, we seek breakdown of the conditions.

$$B\Big/\frac{\partial B}{\partial r} \equiv R_I \ll R_{\text{cyclotron}} \text{ and } \frac{\partial B}{\partial t}\Big/B \equiv \omega_I \ll \omega_{\text{cyclotron}}$$

Taking 2 Mev electrons and measuring distances in earth radii we find on the equator

$$\omega_c = \frac{eB}{\gamma m_0 c} = \frac{1.5 \times 10^6}{r^3} \text{ sec}^{-1}$$

and

$$R_c = \frac{\gamma m_0 c}{eB} v_\perp = \frac{\gamma m_0 c v}{e\sqrt{BB_m}} = 2.7 \times 10^3 [r_m{}^3 r^3]^{1/2} \text{cm}.$$

where r_m is the position of the mirror point.

Now, we really have no measurements in the upper atmosphere of small scale fluctuations in the field. However, it is postulated that the solar wind acting on the earth will set up Helmholz instabilities in the outer reaches of the earth's field and that these fluctuations in the field will be propagated as Alfven waves throughout the field. If this is so, we can take, perhaps, the frequency spectrum of the surface fluctuations as typical of the frequency spectrum of these Alfven waves. To estimate the upper limit of the wave length of these Alfven waves, we use the fastest speed by taking as the ionization density 10^3 protons per cc which has been postulated for the solar corona. This wave length will decrease as the inverse cube of the distance from the center of the dipole while the cyclotron radius of a particle mirrored at the equator will increase as the cube of this distance. As we go up, therefore, there is an altitude above which the cyclotron radius will exceed the Alfven wave length. Above this altitude we would not expect the magnetic moment to be conserved. These altitudes are shown as dotted lines in Figure 7 versus energy for both protons and electrons.

Above these critical altitudes the particle will experience what we may call magnetic scattering off these Alfven fluctuations and will find the magnetic moment either increased or decreased. Now if it is decreased, that means the particle goes to a lower mirror point and it means that the cyclotron radius is reduced and so the magnetic scattering is reduced. On the other hand, if the magnetic moment increases to a higher mirror point, then the scattering is larger.

The statistical average of all of this will be to force the mirror points to lower and lower altitudes and eventually the atmospheric losses will take over. The solid lines we have plotted show the altitudes at the equator for the mirror point at the surface which is the maximum allowable. For the Argus experiment—several Mev electrons with an equatorial altitude of about one earth radii—there is no trouble with magnetic scattering. On the other hand fairly energetic protons would not be expected on this basis to remain in the outer Van Allen belt.

Now, you will notice that the critical altitude is very insensitive to energy. This insensitivity to energy is equivalent to an insensitivity of the assumed frequency spectrum which we have taken. If this idea, this notion, of Alfven waves in

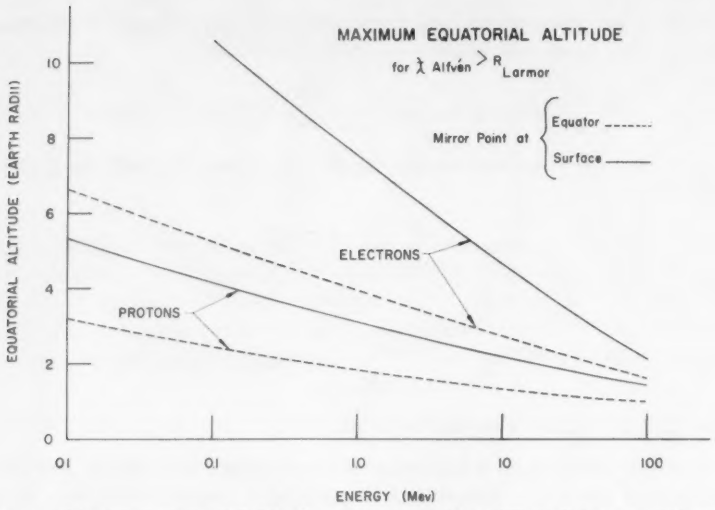

FIG. 7.—Maximum equatorial altitude versus energy for stability of electrons and protons against magnetic scattering.

the upper atmosphere has anything to it then we have a quite effective way of removing protons from the outer Van Allen belts. This notion can then also be connected to the changes in the character of the Van Allen belts associated with magnetic storms, as these are generally associated with large solar winds and large fluctuations in the outer fringes of the field.

Section 7: Effect of Static Magnetic Anomalies.—We have mentioned in Section 3 that one can follow the three dimensional drift of the mirror point by staying on a surface of constant field strength and adjusting the north-south position so as to keep the integral invariant constant. Any particle will have two closed paths, one in the northern hemisphere and one in the southern, traced out by its mirror points. For magnetic latitudes near 45° (that is, for the Argus shell) the northern path is fairly well behaved with north-south excursions of 500 km or so and vertical excursions somewhat less. The southern path is very distorted, however, due to the large area in the South Atlantic where the total field strength is much less than that given by the eccentric dipole (see Fig. 8). This anomaly causes the path to swing south to South America to maintain the integral invariant and to dip very low just west of Capetown to find the surface of constant field strength. This last point is of greatest interest since, as shown in Section 4, it is the average air density around the drift path that determines atmospheric loss rates. Because the air density is exponential, a small downward change in altitude over a fraction of the drift period may contribute overwhelmingly to the average density.

In the case of the Argus data, the situation is more extreme. An electron mirroring over the eastern United States at altitudes below 800 km will find their southern drift paths plunging into the ocean west of Capetown. That is, all the electrons seen by Jason Rockets never lived to drift around the earth again. In passing through the anomaly, some particles at higher altitudes will be scattered

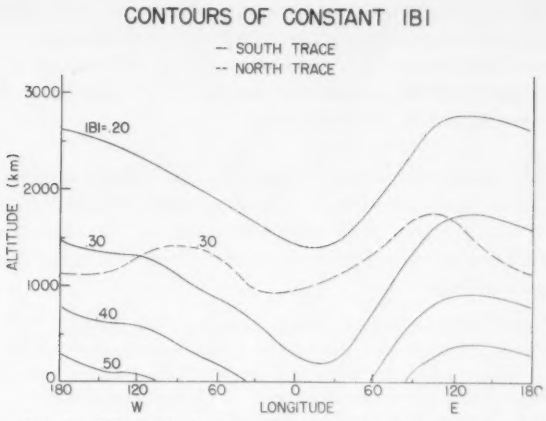

FIG. 8.—The altitude of surfaces of constant scalar magnetic field strength versus longitude at the intersection of these surfaces with the Argus electron shell.

down, however, to repopulate the low altitudes. The only change is from a picture of slow, steady loss to an impulsive loss once each drift period.

The time of transit through the anomaly is about five minutes for a 1 Mev electron. But the characteristic lifetime is just five minutes at 400 km. Now 400 km over the anomaly corresponds to a field strength of about 0.28 gauss. This field strength then is the dividing line between particles that survive passage through the anomaly and those that do not. Inspection of the mirror point density versus field strength curves compiled from rocket and satellite data reveals a sharp change in slope at $B = 0.28$.

A further effect of the anomaly is to increase appreciably the average air density for particles high enough to survive several drift periods. This increased density shortens the characteristic lifetime over the data in Table 2, and may explain

TABLE 2

CHARACTERISTIC LIFETIME FOR 1 MEV ELECTRONS VERSUS ALTITUDE

Altitude h, km	Air Density N, atmos/cc	Scale Height a, km	Characteristic Lifetime for 1 Mev Electron	
			Centered dipole, days	Eccentric dipole where h is minimum altitude, days
0	5.4×10^{19}	10	1.4×10^{-14}*	1.8×10^{-13}*
100	1×10^{14}	20	1.1×10^{-8}	1.0×10^{-7}*
200	2.5×10^{10}	50	7.0×10^{-5}	4.0×10^{-4}
300	3.5×10^{9}	100	7.1×10^{-4}	2.6×10^{-3}
400	1.2×10^{9}	100	2.0×10^{-3}	7.3×10^{-3}
500	4.5×10^{8}	100	5.5×10^{-3}	2.0×10^{-2}
600	1.8×10^{8}	100	1.4×10^{-2}	5.1×10^{-2}
700	6.5×10^{7}	100	3.7×10^{-2}	0.14
800	3×10^{7}	100	8.2×10^{-2}	0.30
900	1×10^{7}	100	0.24	0.88
1000	4.4×10^{6}	100	0.50	1.8
1100	1.7×10^{6}	100	1.3	4.7
1200	8×10^{5}	100	2.8	10.2

* Less than time for one orbit ($\sim 10^{-6}$ days); hence particles with these mirror point altitudes die on their first encounter with the atmosphere.

why the flux was not constant at early times. That is, we may never have made a measurement at any times shorter than a characteristic lifetime.

Having looked at the great effect of the anomaly, we believe now that future calculations should start with the anomaly as the centerpiece and not introduce it as a perturbation as we have done in this paper.

The role of the ambient plasma in checking the charge separation drift was pointed out by Dr. Conrad Longmire. The magnetic invariant formulation for the configuration of the electron shell was suggested by Dr. Theodore Northrup. Dr. Paul Nakada has confirmed the results of the simple scattering theory used here by more rigorous methods.

[1] Alfven, H., *Cosmical Electrodynamics* (Oxford University Press, 1950).
[2] Spitzer, L., *Physics of Fully Ionized Gases* (New York: Interscience Publishers, 1957).
[3] Christofilos, N. C., Univ. of Calif. Rad. Lab. Report, UCRL 5407 November, (1958).
[4] Fermi, E., *Nuclear Physics* (Chicago: University of Chicago Press, 1950).
[5] Harris, I., and R. Jastrow, *Science*, **127**, 471 (1958).
[6] Muehlhause, C. O., and S. Oleska, *Phys. Rev.*, **105**, 1332 (1957).

OPTICAL, ELECTROMAGNETIC, AND SATELLITE OBSERVATIONS OF HIGH ALTITUDE NUCLEAR DETONATIONS—PART I

By Philip Newman

PROPAGATION SCIENCES LABORATORY, ELECTRONICS RESEARCH DIRECTORATE, AIR FORCE CAMBRIDGE RESEARCH CENTER, BEDFORD, MASSACHUSETTS

In the planning of the Argus program it was realized that there might be effects that could be observed on the ground. The Air Force Cambridge Research Center in cooperation with the Stanford Research Institute received the assignment for a setting up of surface instrumentation to make optical and electromagnetic observations. Originally, the chief emphasis was to be placed on the area magnetically conjugate to the detonation area. In the conjugate area some of the trapped electrons after each transit along the magnetic line would enter the atmosphere and possibly cause auroral and associated phenomena. As the electron shell spread around the earth, the conjugate area would move along the geomagnetic latitude. Later the program was enlarged to include the detonation area and a station in Spain to monitor the Explorer IV. It should be stressed that the shortage of time made it impossible to plan more than a minimal program. It was planned to detonate the nuclear bombs so that the conjugate area would be in the vicinity of the Azores. An obvious problem was the possible occurrence of natural geomagnetic disturbances. There was an appreciable disturbance the day of event one and for several days between events two and three.

Possible Effects.—Auroral phenomena: In addition to the luminosity of the aurora which can be observed visually or by cameras in particular the all-sky camera[1] developed by auroral investigators, natural aurora have characteristic spectral lines, the most prominent being the atomic oxygen lines, the red one at 6300A, the green line at 5577, and the violet nitrogen lines at 4278 and 3914.

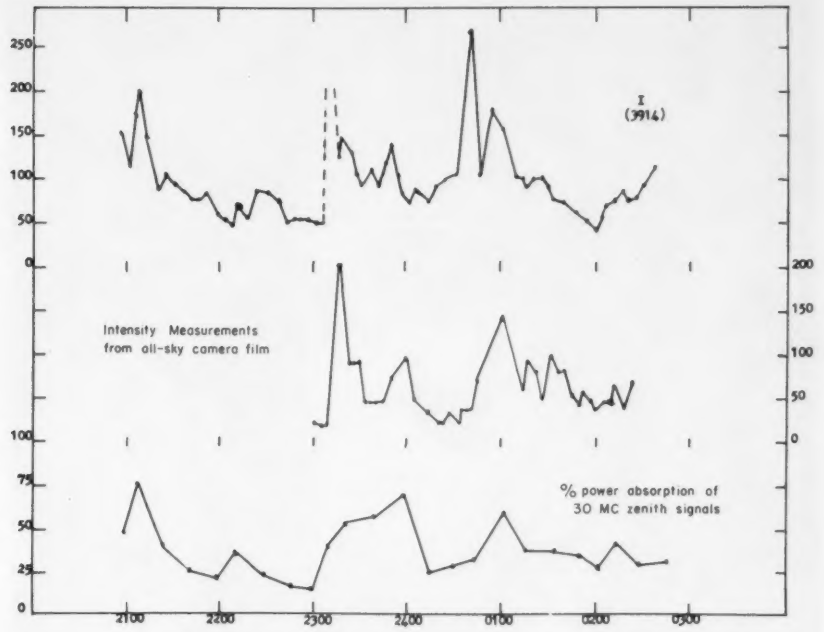

FIG. 1.—Comparison of auroral intensity with zenith absorption.

In addition, there are hydrogen lines, which are attributed to protons. Since an Argus aurora, if any, would be caused primarily by electrons,[2] the oxygen and nitrogen lines would be of principal interest. It was decided to employ the lines 3914, 5577, 6300 and a spectrograph but we were not successful in obtaining the latter two items. The spectrophotometers used were built in AFCRC shops. With increasing height 3914 increases relative to 5577;[3] another important difference between the lines is that in the ever present night airglow at non-auroral latitudes the 3914 line is generally very weak or absent whereas the 5577 line is always present and moreover shows a strong preference for the 100 kilometer level.[13]

During natural aurora the absorption of radio waves in the HF band by the D layer of the ionosphere is increased. This effect can be observed by the diminution in the strength of reception on the ground of cosmic radio noise. Figure 1 shows how these effects are correlated. This was taken from a report of the University of Alaska on investigations sponsored by AFCRC.[1]

A second important property of most auroras is their ability to reflect *radio waves*.[4, 5, 6] This effect is most pronounced in the 10 to 100 mcs range but by employing high power and large antennas, reflections have been obtained up to 800 mc. Figure 2 shows a reflection from an aurora near the city of Quebec, Canada, observed on the scope of an 18 mc radar operated by AFCRC at Plum Island, north of Boston. The lower return is ground black scatter via F layer reflection. Beyond the aurora is ground backscatter via reflection from the aurora or from

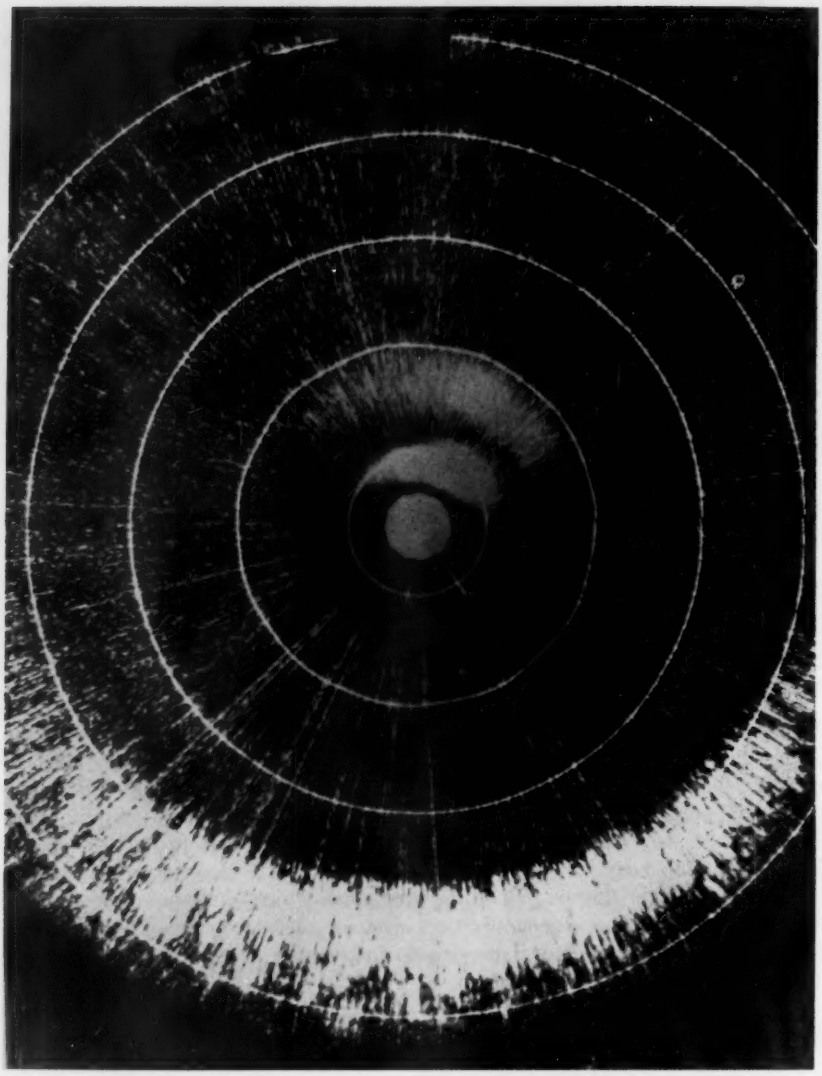

FIG. 2.—18 Mc/s radar return obtained at Plum Island, Mass. Concentric circles indicate 500 mile range. Innermost return is auroral reflection. Ground backscatter reflected via F_2 layer appears at bottom of picture.

Sporadic E accompanying the aurora. The distance to the aurora is about 250 mi. The optimum position for the radar beam is in the direction of geomagnetic north and intersecting the magnetic lines at right angles at ionospheric heights. The frequency selected for the Argus tests was 27 mc.

Other effects which frequently accompany aurora are disturbances in the earth's

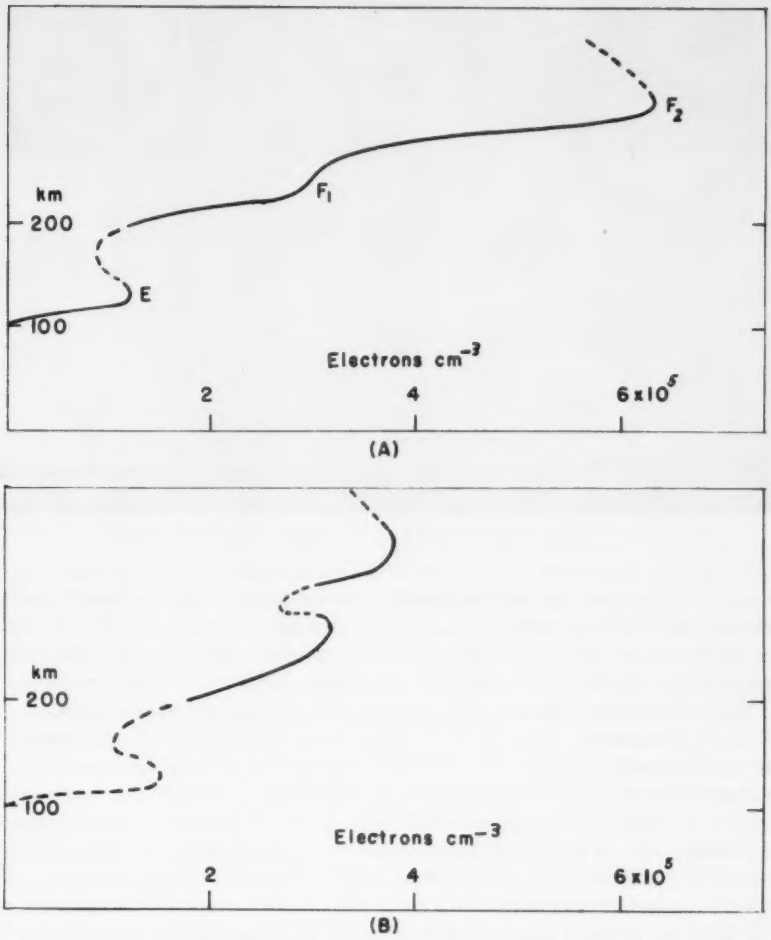

FIG. 3.—Changes in the constitution of the ionosphere from a quiet day (A) to a disturbed day (B).

magnetic field, changes in ionospheric electron densities, formation of Sporadic E, and earth current variations. Figure 3 taken from Harang's "Aurora" shows the change of ionospheric structure during an aurora. Figure 4 shows reflection from a Sporadic E layer, showing also multiple reflections, as obtained on an ionosonde.

Magnetic and electromagnetic effects: The explosion would cause a shock wave in the medium which should propagate as a hydromagnetic wave or Alfven wave, the preferred mode being along the magnetic line of force.[9] This effect should then be detectable on the ground by very low frequency equipment or by magnetometers. A change in the magnetic field could be caused not only by the hydromagnetic wave but also be induced by an electrical current in the atmosphere. Several equipments were available in the Azores region to monitor possible effects. A

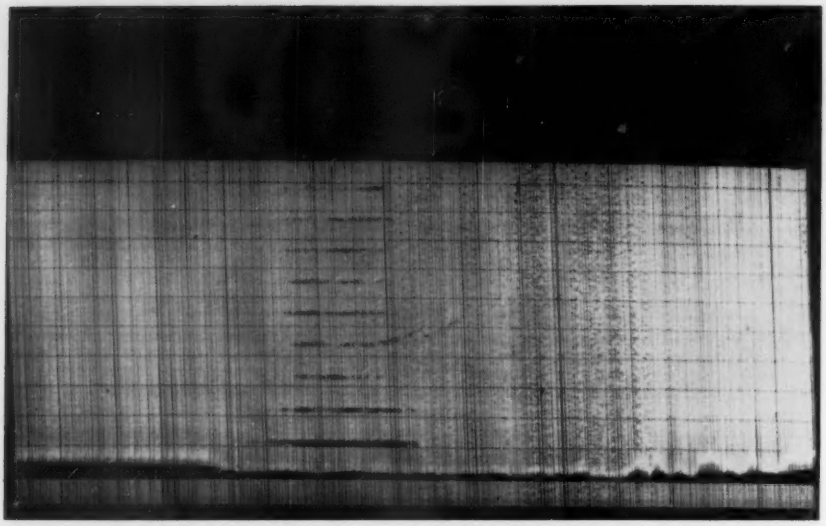

FIG. 4.—Ionosonde records of sporadic E and multiple reflections.

VLF (very low frequency 300 cps to 30 KCS) receiver, was set up with a whip antenna which would measure the electrical field of an electromagnetic effect of the whistler type.[7, 8] This was supplemented with ELF equipment (0.3 to 30 cycles) attached to three mutually orthogonal loops which would measure the rate of change of the magnetic field (Fig. 5). In addition there was a heliflux magnetometer and a variable-mu magnetometer to measure the static magnetic field.

Siting of Equipment.—The VLF, ELF, and magnetic equipment just alluded to were located on the Azores. At the same location there were riometers,[10] which are essentially receivers designed to measure the absorption of the cosmic radio noise. These were at 30, 60, and 120 mc. There were also an all-sky camera and photometers. Two C-97 aircraft, each carrying an ionosonde, were based on the Azores. In addition, one carried an all-sky camera and photometers. The photometers had interference filters 80Å wide at 3914 and 5577 and employed RCA 6810-A. 14-stage photomultiplier. The 27-mc radar for observing radar reflections were on the *U.S.S. Albemarle*. The plan called for this to be sited about 400 miles south of the expected conjugate point. The *Albemarle* also carried an all-sky camera, set of photometers, and a set of riometers. This equipment as did the equipment in the aircraft served as insurance if the conjugate area for any reason did not occur where planned.

At the site in Spain, there was a Microlock receiver to monitor the Explorer IV. In addition, an all-sky camera, photometers, VLF receivers, and riometers were placed there should the auroral effect spread eastward. In addition, regular magnetic stations as well as IGY airglow and auroral stations were available to observe effects.

In the detonation area, there were magnetic detectors, all-sky cameras, and other equipment including VLF receivers, a 30-mc riometer, a 27-mc radar, and photometers at 3914Å and 5577Å.

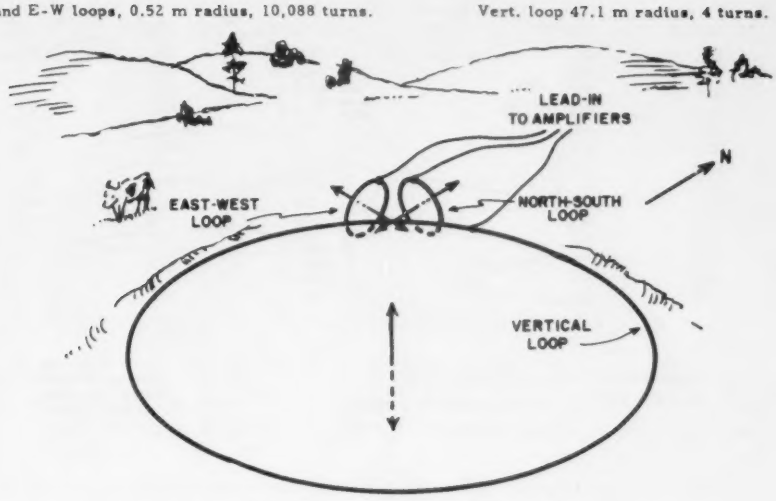

FIG. 5.—Positions of the detector loops used for the extremely low frequency detector. Arrows (normal to plane of loop) indicate direction of maximum sensitivity to changes in H.

Brief Summary.—(*a*) *Visual auroras* were sighted for all events in the launch area; in the conjugate area possibly for the first, and unmistakably for the third. (*b*) *Radar reflections* were obtained for all events except in the conjugate area for number two. (*c*) *Changes in magnetic field* were observed in the launch and conjugate areas; in the conjugate area for events two and three, in the launch area for event three. (*d*) *Very low frequency.* Field strength recordings of GBZ on 19.6 kc made in the Azores and Spain showed decreases of six to twelve db after events two and three. Recordings at Stanford, California, on 18.6 and 17.1 kc showed slight increase. (*e*) Sporadic E was noted in the conjugate area. (*f*) The Explorer IV transmissions were successfully monitored at the Spanish site. The data were incorporated with other satellite data upon which Dr. Van Allen has reported.

A brief discussion of the aurora and ELF results will be given in Part I. In Part II, which follows, the radar, riometer, and satellite results will be covered.

The Visual Aurora in the Congugate Area.—After shot 1 an observer in one of the aircraft reported that he sighted an orange glow in a direction which made sense, twenty-three minutes after the burst. No sighting was reported after the second, although the reported position of one of the aircraft was not too far from the computed conjugate point.

After shot three there was a brilliant auroral display witnessed by Captain Minter and the crew of the *U.S.S. Albemarle* and by civilian scientists aboard. The direction was given as 282° T whereas geomagnetic north was about 340°. The estimated spread in azimuth varied from 3° to 8° to 45° but the later estimate may be the result of reflection from clouds which covered seven tenths of the sky. The elevation was estimated to be from 15° to 30°.

One observer estimated the distance to be about 300–600 miles and the duration

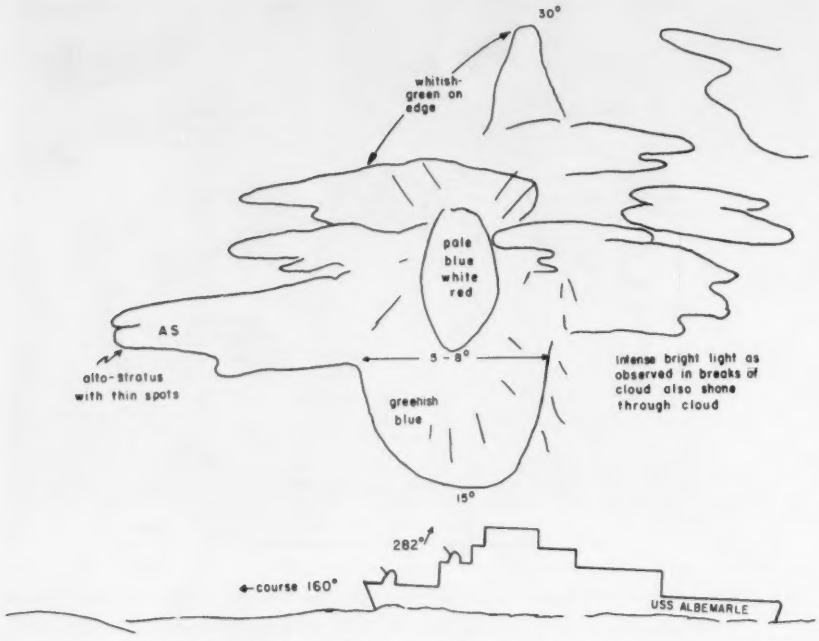

FIG. 6.—Sketch of view from *U. S. S. Albemarle*.

is given as about one-half hour. The aurora observed at the Apia observatory[11] in the South Pacific after the August 1 Johnston Island detonation was reported to have lasted 14 minutes.

The description of one of the observers, Mr. Orange of AFCRC, follows:

"The effect began as a blue-green 'spear' starting close to the horizon, climbing in back of a cloud, and reappearing above the cloud. The effect first appeared about a half a minute after detonation. This is within a few seconds as I had previously set my watch by WWV. A short time after the onset of the effect, about a half a minute, a red crown developed at the head of the bluish spear. The red was distinct but not as bright as the green. For the next minute the red spread out while the blue-green lost intensity. The red aurora deepened in color, began to fade, and after 4 minutes was no longer visible. The blue-green spread out and became an indistinct luminous glow covering about 45 degrees of horizon up to about 30 degrees high. This glow slowly faded and was gone about 32 minutes after the aurora began.

"The brightest part of the initial display was extremely intense as the edges of the cloud which obscured the center of the display were outlined clearly, as if the moon were behind the cloud.

"My position was on the fantail of the *Albemarle* with a clear view of the northern sky. My eyes were adjusted to darkness and I was expecting the event."

The other observers placed the duration at figures varying from 5 to 30 minutes. One observer thinks that the aurora spread from a center and did not observe an

Elevation Angle°	Height, mi	Range, mi	Elevation Angle°	Height, mi	Range, mi
10	60	290	20	60	170
	100	440		100	270
	130	545		130	340
	160	640		160	415
	200	760		200	505
	240	880		300	720
	300	1,000		400	950
15	60	215	30	60	115
	100	325		100	195
	130	440		130	250
	160	525		160	310
	200	615		200	375
	375	1,015		300	500
				400	715

FIG. 7.—Range versus height for various elevation angles.

upshooting pulse, but this he could easily have missed. The observations were complicated by cloud cover which was placed at about seven tenths. In fact, the observations were made possible only by an opportune break in the clouds just before the event.

One of the sketches is presented here (Fig. 6). The top crown of red noted by Mr. Orange would appear to come under Vegard's classification of Type A[12] attributed in natural aurora to proton excitation. It was generally assumed that the effects in the conjugate area, if any, would be produced by electrons from the detonation[13] rather than by heavy ions. In this connection it is worth noting that a recent article by the Russian Galperin[14] gives a high degree of correlation between the occurrence of radar returns and the presence of protons as evidenced by the hydrogen lines in the spectrum of the aurora. Perhaps the movement of heavy ions into the conjugate area should be given more serious consideration.*

It is notable that the duration of the red crown is given as about four minutes. During two of these minutes the radar return was expanding in range, as you will see later. A table can be drawn up (Fig. 7) showing possible heights of the aurora and corresponding ranges for given elevation angles. It will be observed that the estimated distances correspond to auroral heights of about 150–350 miles. The initial radar range of 500 mi gives a height in this bracket, if we assume that the radar return comes from the region in which the visual aurora is taking place. The greater radar ranges would imply either returns from even greater altitudes unless the returns were from non-luminous portions of the sky. A height of 100 km for the aurora at a range of 800 km requires an elevation angle 2° 48′, apparently below the visual aurora. In this case the ground conjugate point is deviated by about 170 km to the northeast of the congugate point for event three as computed at the RAND Corporation by Vestine and Karzas. The range of 800 km and elevation angle of 15° yields a minimum height of about 200 km for the aurora and a deviation of the ground conjugate point of about 280 km to the northwest. Figure 2 shows the possibility of extension of range by ground scatter but this does not appear a satisfactory explanation in view of the smooth extension of radar range reported in Part II, (Fig. 5), but may be applicable to later stages of the radar returns.

If we accept the value of 8° for the width of the visual aurora, and employ the radar range of 500 miles, we find a width of about 75 miles for the visual aurora.

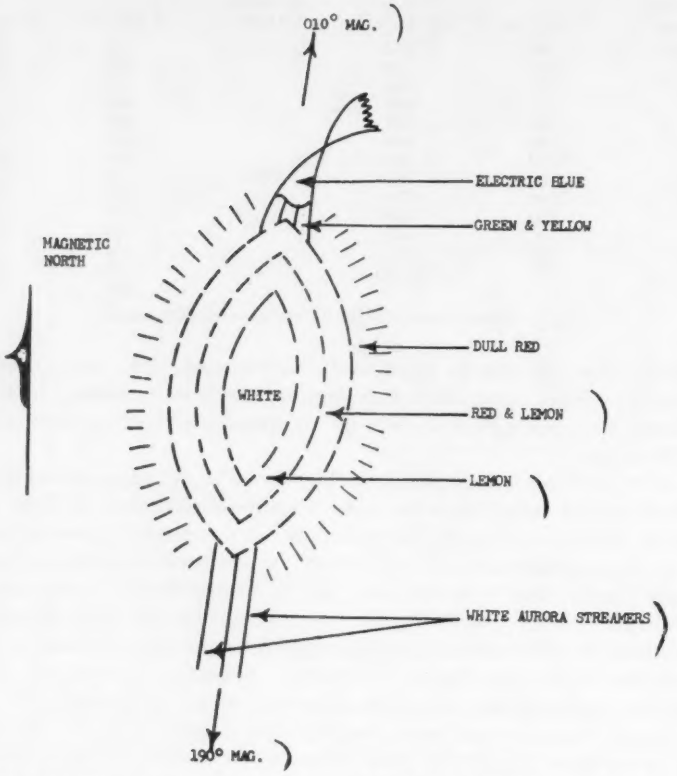

FIG. 8.—Sketch of shot no. 3 phenomena by pilot of aircraft in launch area.

Visual Observations in the Launch Area.—All three events were accompanied by colorful aurora which elongated in a direction making an angle of 10° with the magnetic N-S direction. A sketch of the aurora produced by shot three is shown in Figure 8.

Records were obtained for each of the events of the 5577 and 3914 photometers. These records served as a check on detonation time. We may assume that the effect is mainly caused by the excitation of atmospheric oxygen and nitrogen, since the prompt flash is over in microseconds. The high ratio of 5577 to 3914 (Fig. 9), about 10 to 1 for third event, is much higher than usually found in natural aurora.[15] The result suggests strong excitation of the 100 km level by fission products.

Magnetic Field Effects.—ELF in conjugate area: The ELF receiver operating in the frequency range 0.3 to 30 cps obtained results on events two and three but not on event number one. Two points should be made about *event one*: firstly, it took place during a geomagnetic disturbance which was accompanied by unusually severe radio blackout, and secondly, the magnetic line on which the detonation took place and along which an effect presumably would have been propagated passed through the lower of the recently discovered radiation belts.

For *event two,* two of the coils N-S and vertical were operational. Unmistakable signals were obtained distinguishable from the background by its greater amplitude and lower frequency components. The frequency components of the signal, equated to hormonic equivalent, varied from one-half to 2 cycles per second. The equipment is about twenty times more sensitive at 10 cps than it is at 1 cps so that 10 cps tends to dominate the noise background (Fig. 10). The N-S signal appeared at about plus 13 seconds. Large changes were in evidence for about 13 seconds, with the largest rate of change off scale, occurring at about plus 16 seconds. The V component attained its maximum rate of change about one-half second earlier. Assuming a 1-cp frequency, the N-S magnetic component had a maximum change of about $^1/_2$ gamma, the V component had a much smaller change.

For *event three* (Fig. 11), all equipments operated and the signal went off scale. There may be a signal between plus 4 and plus 6 seconds, but the first large rate of change occurs at about plus 12 seconds on the E-W. The strong signals were not harmonic in nature. Estimates of the amplitudes, assuming a 1 cps pulse, gives the amplitude of the magnetic change as $^3/_4$, $^3/_2$, and $^1/_2$ gammas for N-S, E-W, and V Components.

A sinusoidal waveform signal of 1 cps can be observed plus 31, plus 42 seconds, and possibly at plus 14 all enduring for about 4 to 6 seconds. It was strongest at plus 42 with amplitude of $^1/_4$, $^1/_2$, and $^1/_{25}$ gammas for the N-S, E-W, and V components.

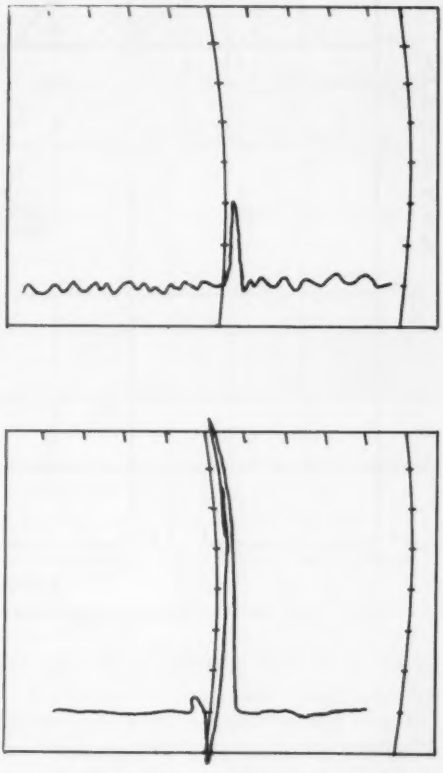

FIG. 9.—Photometer measurements at detonation point. Above, 3,914 Å line, deflection 0.1 volts per division. Below, 5,577 Å, 0.2 volts per division. Chart speed, one large division per second.

The *ground speed* of propagation of the strong disturbance from the detonation point would be about 700 km per second. If we assumed the disturbance followed the magnetic line through the detonation, the average speed would be about 2000 km per second. If we accept a signal at 4 to 6 seconds for event three we find a speed almost three times as great. Recent estimates by Dessler[16] give propagation speeds for an Alfven wave in the upper atmosphere of 300 to 1000 km per second. However, unpublished theoretical work of J. H. Piddington, privately communicated

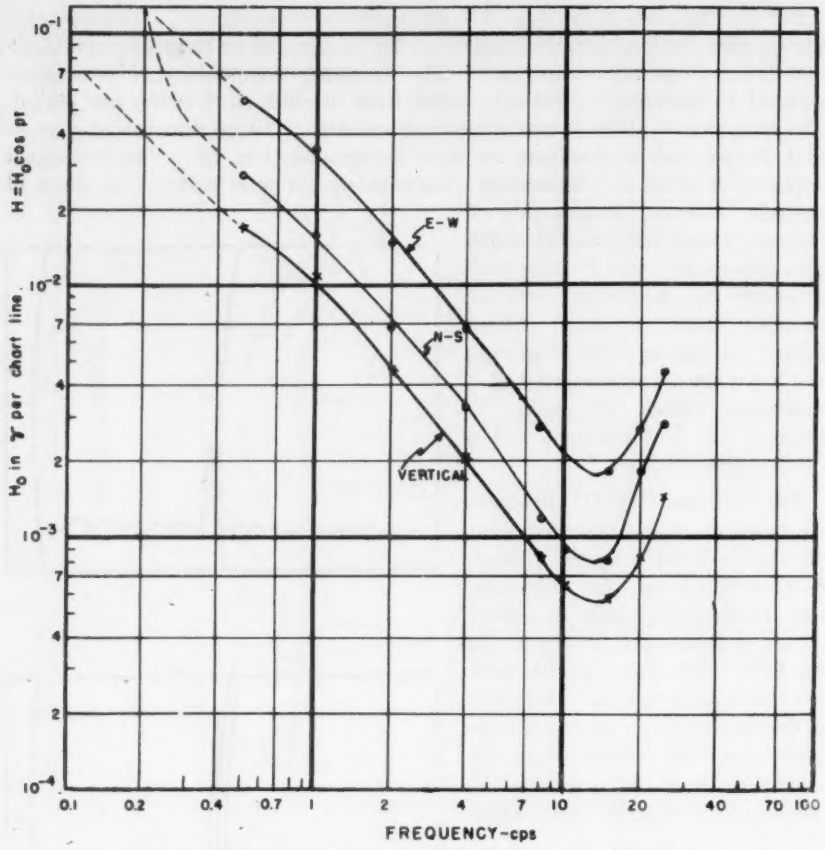

FIG. 10.—Calibration of magnetic detector coils ($1\gamma = 10^{-5}$ gauss).

by G. R. A. Ellis (CSIRO, Australia), indicates the possibility of much higher velocities than those given by Dessler.

It is of interest that a magnetic station operated by the Signal Corps employing a 52 square mile horizontal loop appears to have picked up signals after event three of about $^1/_{400}$ gamma with a periodicity of about 1 cps (Fig. 12). The signals appeared at plus 4 and at plus 20 seconds. Using the plus 4 we find a ground speed of about 3000 km/s and for an upper atmosphere path this would have to be roughly doubled.

The ELF receiver was built at AFCRC for this project. The horizontal loop was 141 meters in diameter and had four turns of wire. The other two, one meter in diameter, had 10,000 turns.

A *magnetic detector* in the launch recorded an initial one-second pulse signal of about 1 cps for event three, with a strength of over 10 gammas followed by longer period pulses. The total duration with irregular fluctuation was about one minute and a half. (Fig. 13).

The *ionosondes* carried in the two C-97's at the Azores were generally not in

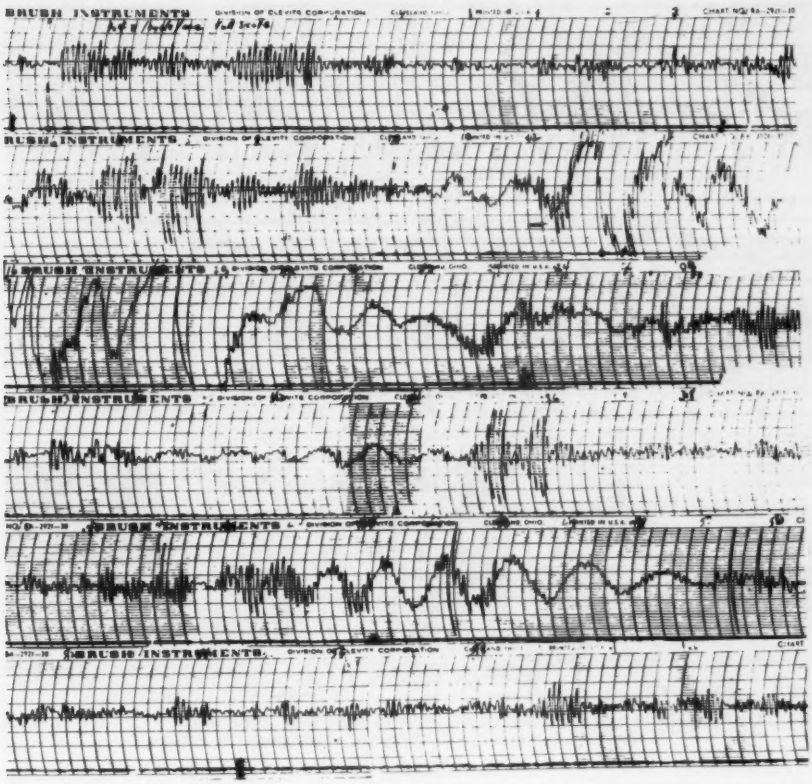

FIG. 11.—ELF magnetic signals received in Azores, September 6, 1958. Chart speed, 5 divisions per second. Arrow indicates detonation. E-W coil.

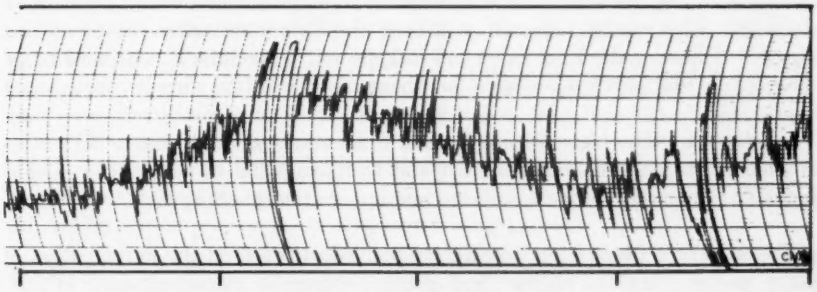

FIG. 12.—1-CPS magnetic bursts. Grand Canyon, September 6, 1958. (Full scale: ±2.5 MV). Lower marks indicate 10 second intervals.

good position to observe ionospheric effects except possibly for event two as mentioned earlier. However, one effect may be worth noting although the direct relation to the event is open to question. For event three, one aircraft because of weather flew east, the other remained on the ground at Lajes. When the ionosonde

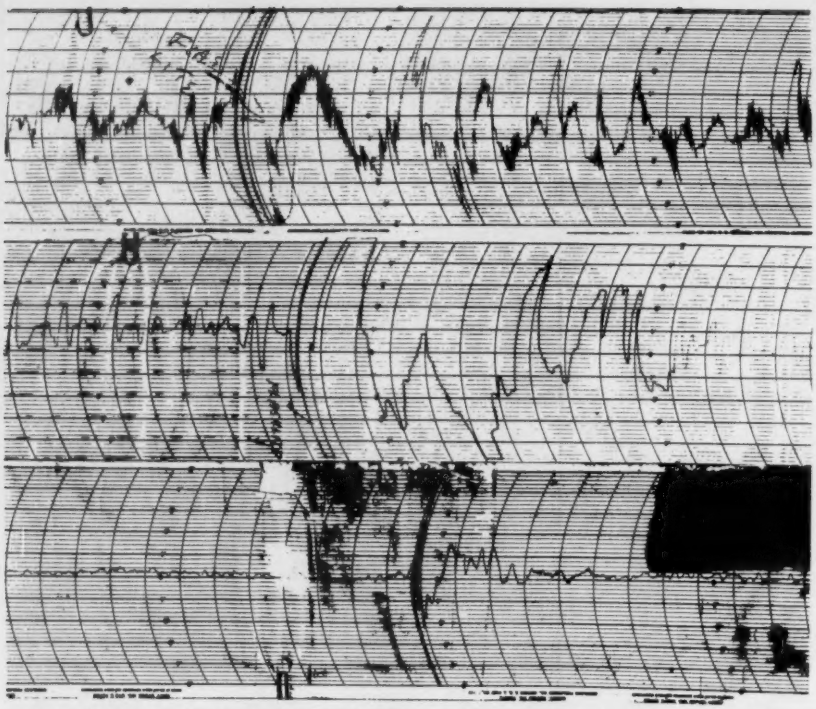

FIG. 13.—Magnetic record at launch point, September 6, 1958. Chart speed: top two rows, 8 divisions per minute; bottom row, 4 divisions per minute.

in the grounded plane was turned on, about 18 minutes after the event it observed Sporadic E. The airborne sounder was about 120 miles east of Lajes and not observing any Sporadic E. On returning the airborne sounder found Sporadic E about 330 miles east of Lajes about 1 hour and 50 minutes after the event (see Fig. 4). The Sporadic E at Lajes had disappeared. Sporadic E continued to be observed by the airborne sounder for about 50 min as the aircraft approached Lajes.

Conclusions.—The high altitude nuclear detonations showed that it is possible to generate auroras of limited extent and hydromagnetic waves; a new tool is available for making controlled auroral and upper atmosphere studies.

The Air Force Cambridge Research Center effort involved the participation of many people. Mr. R. Harvey had general charge of equipment and Mr. N. Oliver and Mr. T. Markham installed the all-sky cameras. Lt. R. Koehler constructed the ELF equipment. Mr. E. Chernosky installed the magnetometers and cooperated in the operation of the ELF equipment and analysis of the records. The advice of E. Maple in the analysis of the ELF records was invaluable. Messrs. Orange, Eng. Straka, and McCabe operated equipment on the naval vessels. Messrs. Miner, Cameron, Shapiro, and Romano operated the ionosondes in the

C–97. The formidable logistics was under the able management of Major C. Wood, Captain R. Yates, Captain E. Angell, and Mr. J. Murphy.

* However, A. Omholt in a paper presented at the May 1959 AGARD Symposium on "Ionization above F max" attributes both high and low auroras to fast primary electrons.

[1] AFCRC TR-56-121 ASTIA #110177

[2] Fowler, P. H., and C. J. Waddington, *Nature*, **182**, 128 (1958).

[3] Harang, L., *The Aurorae* (New York: John Wiley & Sons, 1951), chap. 3.

[4] Bullough, K., and T. R. Kaiser, *Jour. Atm. Terr. Phys.*, **1**, 189 (1945); **2**, 198 (1955); Kaiser, T. R., *The Airglow and the Aurorae*, (New York: Pergamon Press, 1955), p. 156.

[5] Peterson, A., S. Villard, R. Leadabrand, and P. Gallagher, *Jour. Geophys. Res.*, **60**, 497–512 (1955).

[6] "Absorption, Radio Reflections, Ionosonde Techniques," *IGY Annals*, **3**.

[7] *IGY Annals*, **1**.

[8] Gallet, R. M., *Proc. IRE*, **47**, 211 (1959); Helliwell, R., and M. Morgan, *ibid.*, 200; Hines, *Jour. Atm. Terr. Phys.*, **2**, 36, (1957).

[9] Spitzer, L., "Theory of Fully Ionized Gases"; Grad, H., "Propagation of Magnetohydrodynamic Waves," New York University, TID 4500, Jan. 15, 1959.

[10] Little, C. G., and H. Leinbach, *Proc. IRE*, **47**, 315 (1959).

[11] Cullington, A. L., *Nature*, **182**, 1365 (1958).

[12] Vegard, L., in *Threshold of Space*, ed. M. Zelikoff (Pergamon, 1957), p. 22 seq.

[13] Koomen, M. J., *et al.*, in *Threshold of Space*, ed. M. Zelikoff (Pergamon, 1957), p. 222.

[14] Galperin, G. L., *Planetary and Space Science*, **1**, 60 (1959).

[15] Harang, L., *The Aurorae* (New York: John Wiley & Sons, 1951), p. 77.

[16] Dessler, A. J., "Propagation Velocity of World Wide Sudden Commencements of Magnetic Storms," *Jour. Geophys. Res.*, **63**, 405 (1958); *ibid*, **63**, 507, (1958).

OPTICAL, ELECTROMAGNETIC, AND SATELLITE OBSERVATIONS OF HIGH-ALTITUDE NUCLEAR DETONATIONS—PART II

BY ALLEN M. PETERSON

STANFORD RESEARCH INSTITUTE, STANFORD UNIVERSITY

Radar Observations of Auroral Ionization.—It was anticipated that a by-product of the explosion of a nuclear device at very high altitudes might be the generation of an artificial aurora. Not only would a fraction of the high energy β-decay electrons enter orbits which would result in penetration of the ionosphere, but highly ionized heavy-particle debris from the explosion might enter the ionosphere and participate in the generation of an aurora. Since ionized particles are constrained to follow the earth's magnetic field, it was believed that auroral displays might not only be observable in the southern hemisphere near the location of the explosion, but also in the northern hemisphere near the magnetic conjugate point.

In the case of the natural aurora, it has been discovered that the mechanism which creates the visual display also creates ionization[6, 9] in the upper atmosphere which is capable of reflecting and absorbing radio waves. As a result, radio and radar methods of investigating the aurora have within the last two decades become very valuable in research on the aurora. A radar does not see exactly the same thing as the eye or the camera, but it has the great advantage that it can detect auroral ionization through clouds or in daylight. Radio and radar echoes have, in the case of the natural aurora, been observed at radio frequencies between a few megacycles

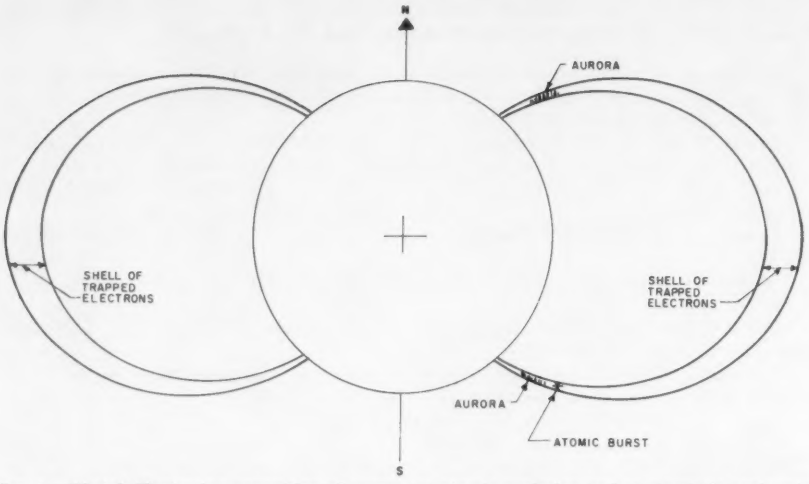

FIG. 1.—Sketch illustrating generation of auroras at atomic explosion and at magnetic conjugate region.

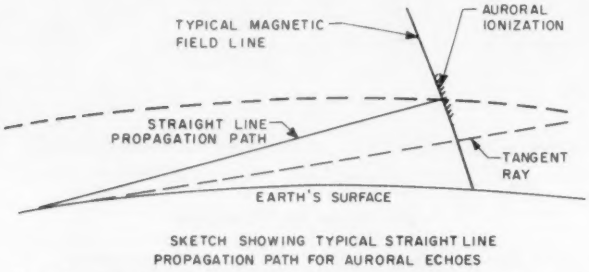
TYPICAL MAGNETIC FIELD LINE
AURORAL IONIZATION
STRAIGHT LINE PROPAGATION PATH
TANGENT RAY
EARTH'S SURFACE
SKETCH SHOWING TYPICAL STRAIGHT LINE PROPAGATION PATH FOR AURORAL ECHOES

FIG. 2.—Sketch illustrating reflection at perpendicular incidence from elongated ionization irregularities aligned with the earth's magnetic field.

and a few hundred megacycles. The number of auroral radar echoes observed and their strength is markedly dependent on the orientation of the earth's magnetic field lines with respect to the observing location. Reflection of radio waves from the aurora behave as if scattering takes place in a region filled with long thin columns of ionization aligned along the earth's magnetic field. Radar echoes are obtained most readily when it is possible to make a perpendicular reflection at an earth's magnetic field line in a region in which auroral ionization is being produced. The requirement for perpendicular incidence is not an absolute one and is a function of frequency. A sketch showing the geometry involved is presented in Figure 2.

The net effect of the aspect sensitive reflection geometry is to limit the region of the sky from which radar echoes are observed. As a result of the geometrical considerations, interpretation of radar echoes is a difficult problem. Not only is the occurrence of the aurora required in the region which the radar is surveying, but also the geometrical conditions must be appropriate.

Figure 3 illustrates the regions in which perpendicular reflection at an earth's

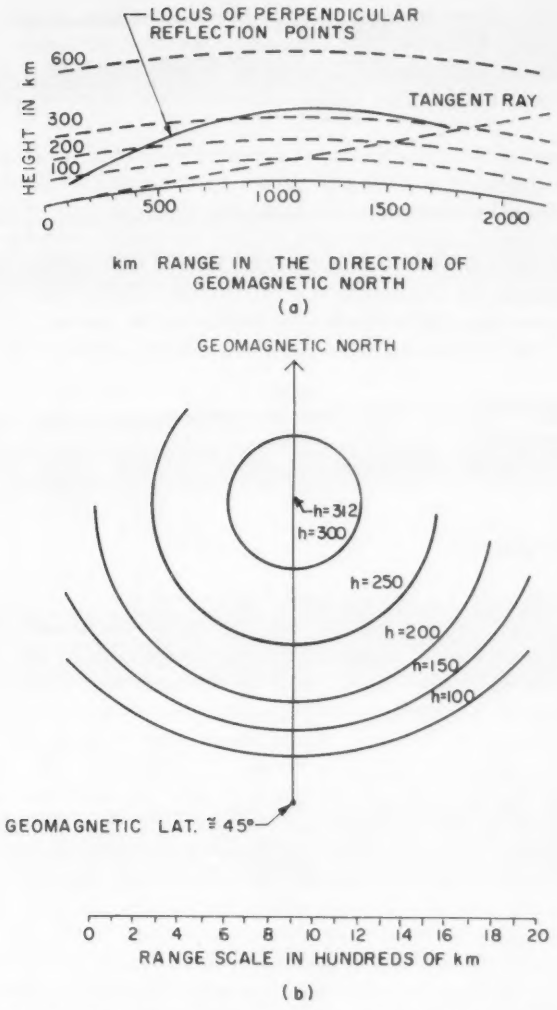

FIG. 3.—Regions at which perpendicular reflection at an earth's magnetic field line can take place.

magnetic field line can take place for a radar location near 45° magnetic latitude. It can be seen in Figure 3a that, looking in the geomagnetic north direction, perpendicularity can be achieved at heights up to about 300 km. In Figure 3b is shown a polar plot of the contours of perpendicular reflection for different heights above the earth. In all cases reflection geometry is favorable only for regions to the north of the radar for a location in the northern hemisphere, such as was assumed in constructing Figure 3. For radar locations in the southern hemisphere, the favorable reflection regions would lie to the south of the radar.

It is known from the natural aurora that the height region of most intense ionization occurs near 100 km and that radar echoes are most often observed from such heights. However, just as visual displays are observed at heights of several hundred kilometers, so too are radar echoes sometimes observed up to about 300 or 400 km.

In the interest of obtaining maximum detection sensitivity for auroras with relatively simple equipment, experimental radars operating at a frequency near 27 Mc were designed and constructed for use in the Argus experiments. Equipment with similar characteristics has been installed in many parts of the world as a part of the United States IGY program in ionospheric physics.[8] These radars (backscatter sounders), operating at frequencies near 30 Mc, have proven very satisfactory in studies of echoes from field-aligned irregularities in the aurora. The characteristics of the 27 Mc radars used in the Argus experiments are shown in Table 1. In

TABLE 1

Frequency	27 Mc
Peak power	7.5 kw
Pulse length	Adjustable 150–1000 μsec
Pulse repetition frequency (PRF)	Adjustable 15–30 pps
Antenna	3 element Yagi array horizontally polarized, mounted for rotation in azimuth
Receiver bandwidth	1–6 kc

the Argus experiments, two of these special radars were used, one on a ship near the location of the explosion, and one on a ship near the conjugate magnetic field point.

Auroral radar echoes were obtained on all three explosions by the radar near the launch end, and on events one and three by the radar near the conjugate points. Negative results on the conjugate end radar for event two were probably caused by unfavorable location of the ship. Range versus time photographs of the echoes observed at the conjugate region radar are shown in Figure 4. The echoes appeared within seconds after the explosion at the launch radar, and somewhat later, although less than a minute, at the conjugate region radar. The echoes were observed in one case for one hour and fifteen minutes at the conjugate region radar, and in one case for five hours at the launch region radar.

In Figure 4 it can be seen that the echoes moved outwards in range. Although initially there is a single well-defined echo, at some stages in the echo development more than one trace is visible, and the echo at times becomes very spread in range. It appears probable that part of the time a sporadic ionization patch existed, which was capable of reflecting the 27 Mc radio waves down to earth at a distant point and which produced "ground scatter" echoes. However, it is believed that most of the observed echoes were directly back-scattered from field-aligned irregularities. A combination of visual auroral observations and the radar records of Figure 4 permitted the location of the conjugate magnetic point to be determined. This measured conjugate point agreed with theoretical expectations to within less than 100 km.

VLF Radio Observations.—Observations of the variations in propagation characteristics of radio waves in the VLF range (300–30,000 kc) have for many years been known to be a sensitive indication of conditions in the lower ionosphere (below 100 km height). In addition to VLF communications transmitters, operated in many

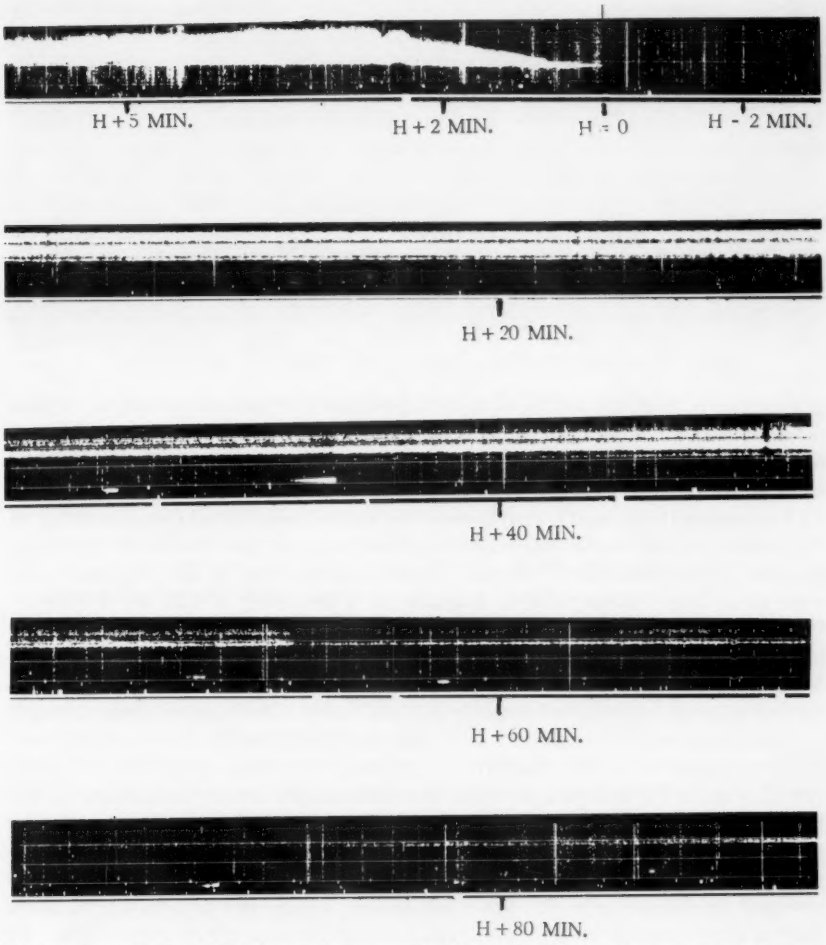

FIG. 4.—Auroral echoes following explosion number three. Radar in region near magnetic conjugate point.

parts of the world, the "atmospherics," signals propagated from lightning strokes, provide a source of signals which can be recorded in the VLF range. The VLF signals can be enhanced by increased ionization at the height of reflection in the ionosphere and reduced by absorption caused by increased ionization at lower levels of the ionosphere through which the signals must pass.

For the Argus experiments, VLF equipment was designed, constructed, and installed at operating locations by personnel of the Stanford University Radio Propagation Laboratory under the direction of Professor R. A. Helliwell. This equipment consisted of a broadband receiver, loop antenna, magnetic tape recorder, and associated calibration equipment and was similar to the IGY whistler recorders.[4]

During the Argus experiments, VLF recordings were made in the launch region,

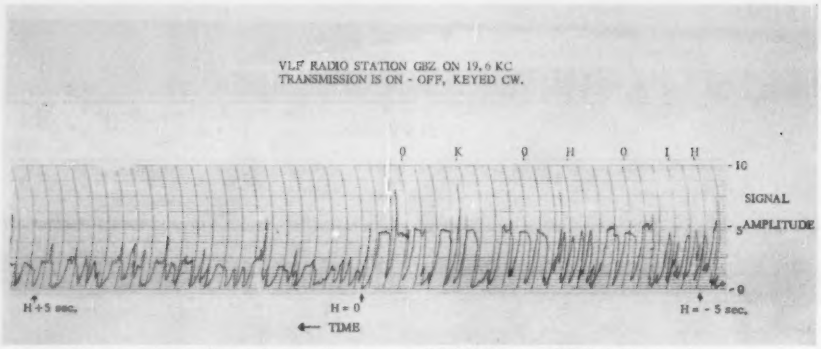

FIG. 5.—Recording taken in the Azores of VLF Station GBZ in England.

at a number of locations in the conjugate region, and at Stanford University. These recorded data have been studied in a search for possible effects on the ionosphere which might result from the trapped shell of electrons which extended around the world or for possible local effects in the launch or conjugate area. The broadband VLF recordings (300–30,000 cps) include the narrow band signal from a number of VLF radio stations which was studied by observing the tape recordings through a tunable narrow-band filter. Signals of particular interest in the conjugate area were those from station GBZ in England on a frequency of 19.6 kc (Figure 5). Recordings of this station taken in the Azores and in Spain showed abrupt decreases in signal strength, following Argus events two and three, which amounted to from 6 db to 12 db. (Event one data have not yet been analyzed.) The signals remained reduced in strength for nearly one-half hour. Decreases in signal strength in Spain commenced a few seconds later than they did in the Azores which are closer to the conjugate area. At Stanford in California, the signals from NPG in Seattle (18.6 kc) and NDT in Japan (17.1 kc) were analyzed for events two and three, and both showed slight increases in signal strength for some minutes following the explosions. A possible explanation for these observations is that the electrons scattered from the Argus shell (and near the conjugate area the ionized debris) actually changed the ionization in the lower ionosphere, apparently penetrating to lower heights and producing absorption in regions relatively near the conjugate area. In the Pacific, far removed from the original injection region in the Atlantic, an enhancement of signal strength was observed, perhaps indicating increases in ionization density near the reflection points.

In addition to these observations, the level of signal strengths from distant atmospherics is being analyzed to seek further substantiation of these results. It had been suggested that whistlers or other VLF noise emissions[2, 4] might accompany the nuclear explosions and might be observable in the launch or conjugate areas. A search of the VLF recordings has been made for signals of this type, but none have been found.

Riometer Measurements.—The riometer[5] is a sensitive, self-balancing, noise measuring piece of equipment which has been designed for routine measurement of ionospheric absorption by monitoring the power level of cosmic noise which has propagated through the ionosphere. Application of this technique during the IGY

has resulted in quantitative measurements of ionospheric absorption in the high frequency band during auroral disturbances and sudden ionospheric disturbances (SID's).

In the riometer a local noise source is continuously made equal to the noise power from the antenna. The receiver is used merely as a detector of inequalities between the noise output of a temperature-limited vacuum noise diode and the noise from the antenna; it generates a D-C output voltage which is proportional to this noise difference. The D-C voltage is used to change the filament temperature and hence the output noise power of the noise diode in such a direction as to reduce the inequality of antenna and noise diode powers to zero. The noise power from the noise diode is directly proportional to the D-C current flowing through it, and, therefore, the antenna noise power can be measured on a linear scale by recording the noise diode current on a pen recorder.

The IGY riometers operate on a single frequency near 30 Mc and have been used with great success in a synoptic network during the IGY. As a part of the Argus measurements, three riometer stations were operating on frequencies near 30, 60, and 120 Mc. The antennas were vertically-directed three-element Yagi arrays of about 60° half-power beamwidth. A single frequency (30 Mc) riometer was operated near launch end during the Argus experiments.

These units all operated satisfactorily during the Argus experiments, although the results were negative in the sense that no significant absorption increases were recorded following the three events.

The riometer measurements demonstrated that although there was sufficient perturbation of the ionosphere by the Argus experiments to produce reduced signal strength or absorption of VLF signals, the perturbations were either not large enough or not widespread enough to be detectable on equipment which is sensitive to absorption in the high frequency range.

Another purpose of the riometer network during the Argus experiments was to provide a sensitive indication of noise generation by trapped electrons. High-energy relativistic-velocity electronics spiralling in magnetic fields have been postulated as a possible mechanism for the generation of cosmic noise in nebulae. In addition, a number of years ago Schwinger[9] calculated the radiation to be expected from high-energy electrons which circulate in strong magnetic fields in the synchrotron particle accelerator. The generation of synchrotron radiation was confirmed by Tomboulian and Hartman[11] at optical frequencies both as to spectra and angular distribution in the ultraviolet frequency range, using the Cornell synchrotron. Based upon the paper by Schwinger and the high-energy particle counting rates measured by Van Allen in the natural radiation belt, calculations have been made by Dyce and Nakada[2] of the possibility of detecting radio frequency radiation from the Van Allen belts. Their calculations show that, assuming the Van Allen belts are of trapped electrons, in the vicinity of 30 Mc the ratio of synchrotron radiation to cosmic radiation from the coldest part of the sky is about 0.4 per cent. The noise ratio is expected to drop off rather rapidly with increased observation frequencies. Thus the synchrotron radiation is not likely to have been observed on existing equipment, but might reasonably be extracted by careful radiometric methods. Dyce and Nakada suggest that, although synchrotron noise from the Van Allen belts and cosmic noise are difficult to resolve because both are wide-band noises and

both are present at all times, they are unlike in the following ways: (1) they have a different power spectrum; (2) synchrotron radiation is mainly radiated normal to the magnetic field lines, whereas cosmic noise arrives from wide angles in the sky; and (3) synchrotron radiation is linearly polarized when viewed normal to the magnetic field, whereas cosmic noise has been experimentally shown to be unpolarized. Antenna directivity and resolution by polarization determination both appear to offer promise for a means to experimentally observe synchrotron radiation from the Van Allen belts.

As a result of the satellite particle counting experiments in the Argus experiments and comparison of electron density in the trapped electron shells with counting rates in the Van Allen belts, it is now believed that the level of synchrotron noise from the Argus experiments should have been below the detectable level with even very refined equipment such as outlined above. The riometer experiments provide confirmation that the noise level was in fact below the level of conventional cosmic noise radiometers even in the 30 Mc frequency range. No detectable increases in noise levels were observed at any time on the riometers which were operated in the conjugate area during the Argus experiments.

[1] Christofilos, N. C., *Trapping and Lifetime of Charged Particles in the Geomagnetic Field*, University of California Radiation Laboratory Report, UCRL 5407 (November, 1958).

[2] Dyce, R. B., and M. P. Nakada, "On the Possibility of Detecting Synchrotron Radiation from High-Energy Electrons in the Van Allen Belts," *Stanford Research Institute Memorandum* (March, 1959).

[3] Gallet, Roger M. "The Very-Low-Frequency Emissions Generated in the Earth's Exosphere," *Proc. IRE*, **47**, 211–232 (1959).

[4] Helliwell, R. A., and M. G. Morgan, "Atmospheric Whistlers," *Proc. IRE*, **47**, 200–209 (1959).

[5] Little, C. G., and H. Leinback, "The Riometer—A Device for the Continuous Measurement of Ionospheric Absorption," *Proc. IRE*, **47**, 315–319 (1959).

[6] Nichols, Benjamin, "Auroral Ionization and Magnetic Disturbances," *Proc. IRE*, **47**, 245–255 (1959).

[7] Nicolet, Marcel, "The Constitution and Composition of the Upper Atmosphere, "*Proc. IRE*, **47**, 142–147 (1959).

[8] Peterson, A. M., R. D. Egan, and D. S. Pratt, "The IGY Three-Frequency Backscatter Sounder," *Proc. IRE*, **47**, 300–315 (1959).

[9] Schwinger, J., "On the Classical Radiation of Accelerated Electrons," *Phys. Rev.*, **75**, 1912 (1949).

[10] Stormer, C., *The Polar Aurora* (Oxford: Clarendon Press, 1955).

[11] Tomboulian, D. H., and P. L. Hartman, "Spectral and Angular Distribution of Ultraviolet Radiation from the 300 Mev Cornell Synchrotron," *Phys. Rev.*, **102**, 1423–1447 (1956).

*ON CHARGE TRANSFER COMPLEXES BETWEEN SUBSTANCES OF BIOCHEMICAL INTEREST**

By Irvin Isenberg and Albert Szent-Györgyi

INSTITUTE FOR MUSCLE RESEARCH AT THE MARINE BIOLOGICAL LABORATORY, WOODS HOLE, MASSACHUSETTS

Communicated June 16, 1959

Introduction.—While a considerable amount of work has been done on charge transfer complexes, the concepts and ideas developed have not, as yet, found much application to problems of biochemical interest.

Spectral evidence indicates that tryptophan forms a charge transfer complex with riboflavin.[1] The work of B. and A. Pullman[2] strengthened the argument that the complexing of tryptophan and riboflavin involved charge transfer by showing that tryptophan was a much better electron donor than the other aromatic amino acids and that riboflavin was a very good electron acceptor. These authors[2,3] calculated the energy levels of a number of biologically significant molecules using molecular orbital methods. These energy levels serve as measures of the effectiveness of the various molecules as electron donor or acceptors. A high energy level of the highest filled orbital means the molecule is a good electron donor while a low energy level of the lowest empty orbital means the molecule is a good acceptor. These energy levels enable one to predict, at least in part, which molecule pairs will complex. This paper will present evidence for the existance of several such complexes, namely,[4] DPNH–FMN, TPNH–FMN, tryptophan–DPN^+, tryptophan–TPN^+ and FMN–FMN. Further evidence that tryptophan and serotonin complex with FMN will also be presented.

Methods.—Complexing is greatly enhanced in aqueous systems by freezing. It is, therefore, convenient to study complex formation by chilling the samples to dry ice temperatures.

Complexes were studied mainly at −78°C, either visually or by means of a Beckman DK–1 Spectrophotometer equipped with an integrating sphere for studying the absorption spectrum of scattering systems.

The integrating sphere was modelled after the system of Strehler and McGee[5] and adapted for use with a Beckman DK–1 spectrophotometer. A double-chambered plastic box was built to fit the Beckman housing. The box was provided with baffles to prevent the direct passage of light. It was then coated with a number of layers of a slurry of MgO and dissolved lucite in ethylene dichloride. Afterwards the box was held in MgO smoke for a final coating.

The system was checked by measuring a variety of known absorption peaks in the liquid state both in the integrating sphere and in the customary way. The position of the peaks did not differ by more than a few millimicrons. It should be noted, however, that the system described here has a significant emission artifact which, in some cases, proves quite useful. Any light emitted by the sample will be collected by the integrator. This will show itself in an apparent lowering of the absorption in any band that leads to emission. This artifact may be demonstrated by studying the absorption of a dye such as thionine. This has two peaks, one at 590 mμ and the other at 560 mμ. At room temperature absorption in the first peak leads to fluorescence while absorption in the second peak does not lead to emission. As a consequence the height of the first peak, relative to the second, is much lower when measured in the integrator than when measured in the usual fashion.

Results.—At −78°C the DPNH–FMN complex, as well as the TPNH–FMN

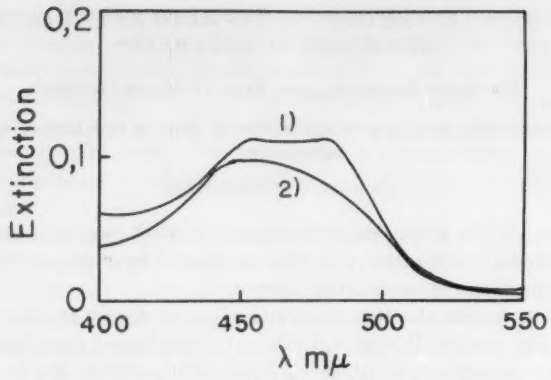

FIG. 1—(1) Absorption of 2×10^{-5} *M* FMN and 2×10^{-5} DPNH. (2) Absorption of 2×10^{-5} *M* FMN. −78°C. in integrating sphere.

complex appears reddish as compared with an FMN control. This in itself demonstrates that complex formation has occurred. In the presence of 10 per cent concentrated hydrochloric acid the color change is marked even at room temperature. In the integrating sphere the absorption spectrum shown in Figure 1 was obtained. It may be seen that, compared to the FMN control, the shoulder at 485 mμ is strongly enhanced. This shoulder appears in the same position as the shoulder that occurs when tryptophan complexes with riboflavin.

The formation of a DPNH–FMN complex may also be demonstrated at room temperature using ordinary absorption methods. This requires higher concentrations of DPNH. Difference spectra of, for example, 10^{-3} M DPNH plus 10^{-4} M FMN versus 10^{-4} M FMN shows a marked drop in the 450 mμ FMN absorption in the complex plus a rise in the absorption at about 500 mμ.

DPNH–FMN complexes have no emission when viewed before a UV lamp.

It should be noticed that the FMN control also shows a shoulder at 485 mμ. It is important to note that FMN, as judged by its energy levels, is not only a good electron acceptor but that it is also a good donor. Consequently we may expect self charge transfer complexes in FMN and these lead to the appearance of a shoulder at 485 mμ.

The self complexing of FMN may be demonstrated in other ways. At room temperature, because of the emission artifact, even FMN concentrations as low as 10^{-4} M will show a distinct shoulder if absorption measurements are made in the integrating sphere. If spectra are run without the integrator much higher concentrations are needed to demonstrate the shoulder but, even here, at concentrations above 10^{-2} M, the shoulder becomes evident.

The complexes of tryptophan and DPN^+ as well as tryptophan–TPN^+ may easily be demonstrated by freezing to dry ice temperatures. At these temperatures the samples appear yellow. In the integrating sphere there is strong absorption in the vicinity of 400 mμ. There is a suggestion that this absorption is composed of at least two very broad peaks, one at about 370 mμ and one at about 420 mμ. The resolution of the instrument in this spectral region is poor when used with the

integrating spheres and consequently the existence of a pair of peaks has not been demonstrated unambiguously.

The tryptophan–DPN$^+$ and the tryptophan–TPN$^+$ complexes show no emission when irradiated by near-U.V.

An interesting case of complexing is furnished by serotonin and riboflavin at acid pH. Upon freezing a sample of 10^{-3} M serotonin, 10^{-3} M FMN and 1 per cent concentrated HCl, the sample turns black. Lower concentrations of serotonin and riboflavin yield gray samples. Such a color change indicates strong complexing that leads to absorption everywhere in the visible. Absorption measurements using the integrator verifies this. There is a broad absorption band at about 570 mμ with a shoulder at 620 mμ. These bands are so broad and so high that they lead to strong absorption almost everywhere in the visible. In addition, there is absorption in the ultraviolet that has a tail extending into the visible. All of these absorptions lead to the striking black appearance of the frozen samples. As shown previously, at neutral reaction the color of the serotonin–FMN complex frozen in dry ice is red-brown instead of black.

A similar examination of tryptophan and riboflavin yields a yellowish green sample rather than a black one. However, absorption measurements show that one is dealing here with the same type of complexing as existed in the case of serotonin but the complexing is weaker. The absorptions at 570 mμ and 620 mμ appear but they are now only shoulders on a strong FMN absorption at 445 mμ.

The above is consistent with the previous observations[2] that serotonin and tryptophan form similar complexes with riboflavin but that serotonin complexes much more strongly.

Conclusion and Summary.—The method of freezing samples affords a simple and effective means for the observation of complexes. Since the procedure is easily reversible one can be confident that a color change indicates complexing rather than an oxidation-reduction chemical reaction. The specificities and the position of the energy levels of the molecules involved lead one to believe that the complexes involve charge transfer. In this way one may demonstrate the existence of charge transfer complexes in the following couples—DPNH–FMN, TPNH–FMN, tryptophan–DPN$^+$, tryptophan–TPN$^+$, FMN–FMN, serotonin–FMN, tryptophan–FMN.

The authors would like to thank Dr. Bernard Strehler for preliminary absorption measurements of dyes in frozen water systems and for assistance in the design of an integrator.

* This research was sponsored by the Grant No. H-2042(C3) of the National Heart Institute, a grant from the Commonwealth Fund, the National Science Foundation, and the United Cerebral Palsy Associations.

[1] Isenberg, I., and A. Szent-Györgyi, these PROCEEDINGS **44**, 857–862 (1958).

[2] Pullman, B., and A. Pullman, *ibid.*, 1197–1202.

[3] *Ibid.*, **45**, 136–144 (1959).

[4] Abbreviations—Reduced diphosphopyridine nucleotide—DPNH, Diphosphopyridine nucleotide—DPN$^+$, Reduced triphosphopyridine nucleotide—TPNH, Triphosphopyridine nucleotide—TPN, Riboflavin $-5'$ phosphate—FMN.

[4] Strehler, B., and M. McGee (in press).

*ON THE ELECTRON PARAMAGNETIC RESONANCE OF ADENOSINE TRIPHOSPHATE**

By Irvin Isenberg and Albert Szent-Györgyi

INSTITUTE FOR MUSCLE RESEARCH AT THE MARINE BIOLOGICAL LABORATORY, WOODS HOLE, MASSACHUSETTS

Communicated June 16, 1959

Recent experiments from this laboratory[1] have suggested that triphosphate compounds might be electron acceptors. Since adenine is a good electron donor[2], the possibility exists that in an adenosine triphosphate molecule the adenine group might donate an electron to the triphosphate. The molecule would then exist in a peculiar internal charge transfer condition, the triphosphate having one electron in excess and the adenine being deficient in one. If such a state is a possible one, then it is likely that the actual state would involve a resonance between two electronic configurations. In one there would be no electron in excess on the triphosphate, while in the other state there would be.

For such a possibility to exist there must be some overlap between the ground state orbital of adenine and an excited state orbital of the triphosphate group. In the extended state of an ATP molecule there can be no overlap. It is therefore of interest to note that, as has been shown by the Courtauld atomic model, folded configurations of ATP may exist in which the triphosphate group is in close proximity to the adenine.[3] In one of these configurations it may be shown that the triphosphate group fits very neatly on the face of the adenine ring structure. In this folded configuration there may be appreciable overlap between the orbitals of the adenine and the triphosphate moieties.

Thanks to the kindness of Professor Hartmut Kallmann and Dr. Neil Wotherspoon of the Institute of Mathematical Sciences at New York University, we were able to make some preliminary studies of ATP using the techniques of electron paramagnetic resonance. The electron paramagnetic resonance assembly was a Varian model, V-4500, equipped with a Varian six-inch magnet. Approximately 50 mg of crystalline ATP (Disodium Salt, Pabst) was placed in a Pyrex test tube of 1.5 mm inside diameter. The tube was evacuated and sealed. The tube was then placed in the microwave cavity and a spectrum run.

A low signal of complex shape was obtained. This signal is shown in Figure 1 along with the spectrum obtained from ADP prepared in a similar fashion. It may be seen that the spectrum of ADP is quite low and is a simple one showing no structure. This is important because atomic models demonstrate that in ADP the phosphates cannot be fitted nearly as well onto the adenine face as in ATP. Furthermore, diphosphate does not have the acceptor properties of triphosphate.[1]

The ATP signal was integrated and compared with an integrated signal from DPPH (diphosphopicryl hydrazyl) which contains one unpaired electron per molecule. On an equivalent molecular basis the ATP signal was lower than that of DPPH by a factor of 10^{-4}.

Several possibilities present themselves for understanding the weakness of the ATP signal.

(1) In a charge transfer complex, the fact that there is resonance between a charge transfer state and a no-charge transfer state means that the spins will

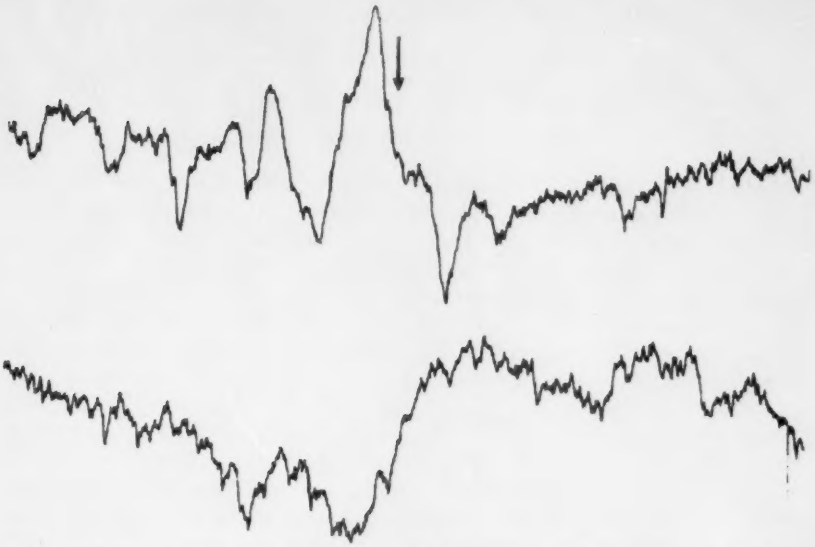

FIG. 1.—First derivative of electron paramagnetic resonance absorption. Upper curve is ATP. Lower curve is ADP. The arrow indicates the free electron resonance. The figure covers 800 gauss.

tend to remain parallel even in the charge transfer state. Only if one of the spins flips due to spin-orbit coupling or some other reason will the complex exist in a state in which e.p.r. absorption is possible. Thus, even if there were quite appreciable charge transfer there might still be relatively low e.p.r. absorption.

(2) In a dry ATP powder only a small fraction of the ATP molecules may be in the folded configuration and only molecules in the folded configuration may show an internal charge transfer.

(3) Even if a relatively large fraction of the ATP molecules were folded in a dry powder the resonance might be such that the contribution of the internal charge transfer state might be small compared to the no-charge transfer state.

(4) A combination of the above might be operative.

(5) The signal may be due to a transition metal impurity in the form of a metal-ATP complex. Judging from the procedure for the manufacture of ATP,[4] such an impurity would be quite a large one. This possibility, though not likely, cannot be ruled out at present.

It is clear that further work must be done on this problem. However, in view of the importance of ATP in cellular metabolism the authors felt that this preliminary note should be published at this time.

* This research was sponsored by the Grant No. H-2042(C3) of the National Heart Institute, a grant from the Commonwealth Fund, the National Science Foundation, the Muscular Dystrophy Associations of America, the American Heart Association, and the United Cerebral Palsy Associations.

[1] These PROCEEDINGS, (to be published soon).

[2] Pullman, Bernard, and Alberte Pullman, these PROCEEDINGS, **44**, 1197 (1958).

[3] Szent-Györgyi, Albert, *Bioenergetics* (Academic Press, 1957).

[4] United States Patent No. 2,700,038. Pabst Brewing Company.

*ENZYMES IN A GLYCOGEN STORAGE MYOPATHY**

By Joseph Larner and Carlos Villar-Palasi

DEPARTMENT OF PHARMACOLOGY, SCHOOL OF MEDICINE, WESTERN RESERVE UNIVERSITY

Communicated by Harland G. Wood, June 22, 1959

McArdle has reported a "myopathy due to a defect in muscle glycogen breakdown"[1] which appears to be a type of glycogen storage disease distinct from the previously reported classes.[2] Schmid and Mahler[3] using muscle homogenates from a patient similar to the one described by McArdle have found that lactate is formed from added glucose 1-phosphate at a rate comparable to that of controls, but addition of muscle glycogen did not increase production of lactate. The content of glycogen in the muscles of this patient ranged from 2.4 to 3.0 per cent. Estimation of phosphorylase indicated a markedly reduced activity of this enzyme.

Recently it has become apparent[4-8] that there is a uridine linked pathway of glycogen synthesis present in animal tissues which involves two enzymes UDPG† pyrophosphorylase, and UDPG glycogen transglucosylase. It was therefore of interest to determine the enzymes of the uridine linked pathway in the muscle of a patient with McArdle's disease. Schmid, Robbins, Traut, and Lipmann[9] have already estimated levels of phosphorylase, phosphoglucomutase, and UDPG glycogen transglucosylase. We report here further results with a sample of the same muscle biopsy, including confirmation of the extremely low phosphorylase activity, and in addition normal activities of UDPG pyrophosphorylase, amylo-1,6-glucosidase, and phosphoglucomutase.

Materials and Methods.—The biopsy of the patient with the disease was obtained from the back. Two control biopsies of rectus and pectoral muscles were obtained during surgical procedures.‡ Homogenates of the rapidly frozen tissue (1:10 w/v) were made in 0.05 *M* Tris 0.005 *M* EDTA pH 8.

Amylo-1,6-glucosidase was determined by the method of Hers (10), using phosphorylase limit dextrin instead of glycogen. UDPG pyrophosphorylase, phosphoglucomutase, and phosphorylase were determined spectrophotometrically (see Table I), using extracts prepared from homogenates by centrifugation at 3000 × *g* for 10 minutes. Activities have been expressed in the units used by Hers.[10]

Results and Comments.—A phosphorylase activity corresponding to 0.23 per cent of the control value was found in the diseased muscle (Table 1). No significant differences in the activities of amylo-1,6-glucosidase, phosphoglucomutase, or UDPG pyrophosphorylase were noted when compared to controls.

It appears that the increased muscle glycogen content in this disease is associated with a markedly diminished phosphorylase activity. From the present and previous data[9] it is apparent that the two enzymes of the uridine pathway are present in the diseased muscle. There appears to be no deficiency in phosphoglucomutase or amylo-1,6-glucosidase activities.

Previous experiments have shown that activation of phosphorylase by epinephrine or glucagon is associated with glycogen breakdown.[11] Glycogen synthesis stimulated by insulin in rat diaphragm in short time experiments takes place with a ratio of inorganic phosphate to glucose 1-phosphate of 305.[7] Activation of phosphorylase in muscle homogenates by addition of phosphorylase kinase results in

TABLE 1

ENZYME ACTIVITIES IN MUSCLE

Patient*	Amylo-1,6-glucosidase Activity, per 1000/mg/hr	Phosphorylase† Activity, μm/g/min	UDPG‡ Pyrophosphorylase Activity, μm/g/min	Mutase§ Activity, μm/g/min
1. McArdle's disease	.006	0.025	28.6	63.4
2. Control	.004	9.20	32.0	67.5
3. Control	.010	10.65	26.3	56.4

* Patient 1, age 52, back muscle; patient 2, age 54, rectus muscle; patient 3, age 55, pectoral muscle.

† Reaction mixture contained (final concentration): inorganic phosphate buffer pH 7.2 1.7×10^{-2} *M*; Mg^{++} 1.7×10^{-2} *M*; TPN, 1×10^{-2} *M*; glucose-1,6-diphosphate 1.7×10^{-6} *M*; phosphoglucomutase (freshly dialyzed) 0.05 ml; AMP 8×10^{-4} *M*; glucose 6-phosphate dehydrogenase (5 mg./ml) 0.01 ml; reaction started by addition of rabbit muscle glycogen, 1 mg.

‡ Reaction mixture contained (final concentration): glycylglycine buffer pH 7.45 2.5×10^{-2} *M*; Mg^{++} 2×10^{-2} *M*; TPN 1×10^{-2} *M*; glucose-1,6-diphosphate 2×10^{-6} *M*; glucose 6-phosphate dehydrogenase (5 mg/ml) 0.01 ml; phosphoglucomutase (freshly dialyzed) 0.05 ml; UDPG 4×10^{-4} *M*; reaction started by addition of pyrophosphate 2×10^{-2} *M*.

§ Reaction mixture as for UDPG pyrophosphorylase except that glucose 1-phosphate 2×10^{-2} *M* was used in place of UDPG and pyrophosphate, and phosphoglucomutase omitted.

glycogen breakdown.[6] These previous findings together with the results reported here may best be explained on the basis of a cyclic mechanism of glycogen synthesis and degradation in which the uridine linked pathway catalyzes the synthesis and phosphorylase the degradation of glycogen.

We wish to thank Mr. Norman E. Brown for technical assistance in these studies.

* Supported in part by a grant from the U. S. Public Health Service A–2366 and from the Cleveland Diabetic Fund and the Fund for the Study of Diabetes and Related Metabolic Disorders.

† The following abbreviations have been used: UDPG, uridine diphosphate glucose; Tris, tris (hydroximethyl) aminomethane; EDTA, ethylene diamine tetra-acetic acid; AMP, adenylic acid; TPN, triphosphopyridine nucleotide.

‡ The muscle biopsy from the patient was kindly supplied by Drs. R. Schmid and P. W. Robbins. The control biopsies were obtained locally through the cooperation of Dr. W. R. Drucker.

[1] McArdle, B., *Clin. Sci.*, **10,** 13 (1951)

[2] Cori, G. T., *Polysaccharides in Biology, Transactions of Third Conference Josiah Macy Foundation* (1957), 124.

[3] Schmid, R., and R Mahler, *J. Clin. Invest.*, **38,** 1040 (1959).

[4] Leloir, L. F., and C. E. Cardini, *J. Amer. Chem. Soc.*, **79,** 6340 (1957).

[5] Villar–Palasi, C., and J. Larner, *Biochem. et Biophys. Acta.* **30,** 449 (1958).

[6] Robbins, P. W., R. B. Traut, and F. Lipmann, these PROCEEDINGS, **45,** 6 (1959).

[7] Larner, J., C. Villar-Palasi, and D. J. Richman, *Proc. New York Acad. Sci.* (in press).

[8] Leloir, L. F., J. M. Olavarria, S. H. Goldemberg, and H. Carminatti, *Arch. Biochem. et Biophys.* **81,** 508 (1959).

[9] Schmid, R., P. W. Robbins, and R. B. Traut, these PROCEEDINGS, **45,** 1236 (1959).

[10] Hers, H. G., *Rev. Int. d'Hepatologie*, **9,** 35 (1959).

[11] Sutherland, E. W., *Phosphorus Metabolism II*, ed. W. D. McElroy and B. Glass (Baltimore: Johns Hopkins Press, 1952), p. 577.

*GLYCOGEN SYNTHESIS IN MUSCLE LACKING PHOSPHORYLASE**

RUDI SCHMID, PHILLIPS W. ROBBINS, AND ROBERT R. TRAUT

THE THORNDIKE MEMORIAL LABORATORY, HARVARD MEDICAL SCHOOL, AND
THE ROCKEFELLER INSTITUTE

Communicated by Fritz Lipmann, June 15, 1959

Mounting evidence for the catabolic rather than the anabolic function of phosphorylase has been forthcoming lately.[1–4] The discovery by Leloir and Cardini[5] of a UDPG-linked glycogen synthesis, on the other hand, presented a most suitable anabolic system where, in contrast to phosphorylase, glycogen synthesis is virtually a one-way reaction. Nevertheless, to find an unambiguous method of assigning intracellular glycogen synthesis to one or the other enzyme appeared to present considerable difficulties. At this juncture, therefore, it was particularly fortunate when a human myopathy was discovered[6] where muscle biopsy showed that a rather high glycogen content was coupled with very low levels of phosphorylase. It seemed probable that the UDPG-linked enzyme was present and that it was responsible for the high glycogen content. We wish to report here on muscle biopsy assays from such a patient which showed the level of the UDPG-linked glycogen synthesizing enzyme to be within the range of normality, in contrast to the again confirmed absence of phosphorylase.

Materials and Methods.—Uridine diphosphate glucose and triphosphopyridine nucleotide were purchased from the Sigma Chemical Company; and 5′-adenosine monophosphate and glucose-1-phosphate were obtained from Schwarz Laboratories. Rabbit liver glycogen was purchased from Mann Research Laboratories and glucose-6-phosphate dehydrogenase, free of phosphorylase and phosphoglucomutase, was obtained from C. F. Boehringer and Sons. This enzyme was dialyzed overnight against 2 mM Tris, pH 7.5 and stored, frozen, in small batches. Under these conditions, it was found to retain its activity for several weeks. Phosphoglucomutase was prepared from rabbit muscle by the method of Najjar[7] and phosphorylase *b* was prepared by the method of Fischer and Krebs.[8] Labeled UDPG was prepared as described previously.[4]

Muscle biopsies were obtained under local or general anaesthesia and were immediately chilled in ice. The tissue was weighed, minced, and suspended in 10 volumes of ice-cold distilled water. Disruption was carried out in a small glass Potter-Elvehjem homogenizer. The homogenates were centrifuged for 10 minutes at 10,000 x g and the enzymatic activities in the supernatant fluid determined. Protein concentration was measured by the method of Lowry *et al.*[9] and the specific activities of all enzymes are defined in terms of protein concentration. The enzymatic activities were determined as soon as possible after homogenization. Aliquots of the homogenate were diluted with glycerophosphate-cysteine, pH 6 (Cori phosphorylase assay), 0.03 *M* cysteine (phosphorylase kinase), or cold water immediately before being assayed.

Phosphorylase was determined by the method of Illingworth and Cori[10] in the presence of AMP at pH 6. The units are those found at pH 6 and are not corrected to the activity expected at pH 6.8. When the homogenate obtained from the patient's muscle was being assayed, 0.2 ml. of undiluted homogenate and 0.2 ml. of

double-strength glycerophosphate-cysteine were used in place of diluted enzyme. It was demonstrated that even under these conditions only about 15 per cent of the glucose-1-phosphate had been converted to glucose-6-phosphate at the end of the incubation period.[10] Phosphorylase was also determined by the method of Wu and Racker.[11] This assay involved a measurement of the rate of TPN reduction when a limiting concentration of phosphorylase was incubated with excess glycogen, inorganic phosphate, phosphoglucomutase, and glucose-6-phosphate dehydrogenase. The incubation mixture contained 50 mM Tris, pH 7.5, 10 mM cysteine, 10 mM sodium phosphate, pH 7.5, 2 mg per ml of glycogen, 0.2 mM TPN, 10 mM $MgCl_2$, 1 mM AMP, phosphoglucomutase, and dialyzed glucose-6-phosphate dehydrogenase. Determinations with crystalline phosphorylase *b* showed that the rate of TPNH formation was directly proportional to phosphorylase concentration over a wide range. Homogenates were diluted with water for this assay.

Phosphorylase kinase assays were carried out according to Krebs and Fischer[12] except that Mg-ATP was used in place of Mn-ATP. Crystalline rabbit muscle phosphorylase, itself free of kinase activity, was used as the substrate. The units are those of Krebs and Fischer, one unit representing the conversion of 100 units of phosphorylase *b* to *a* in 5 minutes at 30°.

Phosphoglucomutase was measured by the method of Najjar[7] and also by a modification of the spectrophotometric method described above for phosphorylase. In this modification, the inorganic phosphate and phosphoglucomutase were omitted and the glycogen was replaced with 1 mM glucose-1-phosphate. It was shown that the rate of TPN reduction was proportional to phosphoglucomutase concentration.

The UDPG-glycogen synthetase activity was determined with glucose-C^{14}-labeled UDPG by a modification of the method described previously[4] (Table 3). The incubation medium was modified by the addition of 0.4 mM glucose-6-phosphate which was shown by Leloir *et al.*[13] to increase the rate of the reaction. The incorporation was determined by the glycogen precipitation method used previously. It was also determined by stopping the reaction with 0.5 ml. of 0.05 *M* acetic acid, adsorbing the nucleotides with about 20 mg. of Norite A, and plating and counting the charcoal supernatant fluid. The two methods gave the same results for control homogenates and for the homogenate obtained from the patient's muscle.

Results.—The levels of phosphorylase and UDPG-glycogen synthetase in three control biopsies and in the biopsy obtained from the patient are summarized in Table 1. In confirmation of the findings of Dreyfus *et al.*,[14] it was found that normal human muscle contains $^1/_4$ to $^1/_3$ the amount of phosphorylase activity found in rat or rabbit muscle. It is possible that the apparent trace of phosphorylase activity found in the patient's muscle may not be phosphorylase at all, but simply a reflection of nonspecific phosphatase activity in the case of the Cori assay, and the formation of traces of glucose-6-phosphate from sources other than glycogen in the spectrophotometric assay. This possibility was strengthened by the finding that the omission of AMP caused almost no change in the amount of enzyme measured. The data available indicate, to say the least, that the amount of activity was less than 0.5 per cent of normal. Mixing experiments have shown that the extract from the patient's muscle did not contain a phosphorylase inhibitor. Dialysis of the extract overnight against 5 mM Tris, 1 mM EDTA pH 7.5, did not result in the appearance of activity.

TABLE 1*

ENZYMATIC ASSAYS ON MUSCLE BIOPSIES

	Patient	Controls (3) Average	Range
Phosphorylase, Cori assay, units/mg protein	0.09	22	16–26
Phosphorylase, spectrophotometric, mμmoles/min/mg protein	0.7	250	170–300
UDPG-glycogen synthetase, mμmoles/min/mg protein	10	21	19–22
Phosphoglucomutase, Najjar assay, milli units/mg protein	30	90	70–100
Phosphoglucomutase, spectrophotometric, mμmoles/min/mg protein	36	150	120–190
Phosphorylase kinase units/mg protein	31	28	21–32

* Muscle biopsies were chilled, weighed, and homogenized in 10 vols. of ice-cold distilled water. After centrifugation, the enzymatic activities were determined by the methods described in detail in the text.

The UDPG-glycogen synthetase is present in normal controls at about the same level found in extracts prepared from rabbit or lamb muscle.[15]

As first shown by Leloir *et al.*,[13] glucose-6-phosphate has an unexplained stimulatory effect on UDPG-glycogen synthetase. As seen in Table 2, glucose-6-phosphate appears necessary in assays of the synthetase even in the crude system. Experiments to be reported elsewhere show that in purified preparations, glucose-6-phosphate appears to be an obligatory component of the system. The patient's muscle tissue contained about $^1/_2$ the amount of enzyme in the controls. The seemingly lower value should reflect a degeneration of the tissue. As seen in Table 1, the extract from the patient's muscle tissue also contained only $^1/_4$ to $^1/_3$ the phosphoglucomutase of control preparations. The damaged state of the muscle in this patient is described in detail by Schmid and Mahler.[16] Dreyfus *et al.*[14, 17] have reported that diseased or damaged muscle contains, in general, lower levels of the glycolytic enzymes.

TABLE 2*

UDPG-GLYCOGEN SYNTHETASE ACTIVITY IN THE PRESENCE AND ABSENCE OF GLUCOSE-6-PHOSPHATE

	Glucose Incorporation into Glycogen cpm/15 min/mg Protein	
Origin of Muscle	No Addition	G-6-P, 4×10^{-4} *M*
Normal 1	1390	4000
Normal 2	273	2280
Patient	571	1870

* The homogenates were prepared and assayed in the manner described in the text.

Measurement of phosphorylase kinase activity in controls gave an average specific activity about $^1/_2$ as great as that reported for crude rabbit muscle extracts by Krebs and Fischer.[12] This lower activity is not surprising, especially in view of the fact that rabbit muscle phosphorylase *b* was used as the substrate rather than human muscle phosphorylase *b*. Surprisingly, the extract from the patient's muscle appeared to have about the same activity as controls. Analogous results were obtained by Nirenberg[18] who found that HeLa cells, although almost devoid of phosphorylase, contained an active phosphorylase kinase system. These findings seem to show the need for an investigation of the distribution and specificity of phosphorylase kinase activity. In a single experiment, it was shown that normal human serum was free of phosphorylase kinase activity.

Discussion.—The abundance of glycogen in these dystrophic muscles in the virtual absence of phosphorylase rather unambiguously connects *in vivo* glycogen synthesis to UDPG-glycogen synthetase. This enzyme was found in the dystrophic muscles in approximately normal amounts when related to functioning muscle tissue. Judging from the symptomatology of the patient, it would seem that only skeletal muscle phosphorylase is lacking. Whether this defect is hereditary and what its relationship to glycogen storage diseases may be deserves further consideration, as does the question of whether the deficiency is caused by an absence of phosphorylase or by the presence of a defective and inactive protein.

When reflecting on the difference between catabolic and anabolic pathways, a more general rule seems slowly to emerge. Development in the area of glycogen synthesis and breakdown runs quite parallel to what is happening in fatty acid synthesis and breakdown. In both cases, metabolism and anabolism were attributed for some time to a chain of reversible reactions proceeding in either direction depending on conditions. It seems now that the anabolic process takes place by a more or less irreversible pathway which may require a rather large energy expenditure. Metabolic energy, it would appear, is treated as a relatively abundant commodity and cellular mechanisms appear not to be designed to save energy.

Summary.—Muscle biopsies have been obtained from normal human subjects and from a patient with a progressive myopathy whose muscles lack phosphorylase. The phosphorylase level in a homogenate of the patient's muscle was found to be less than 0.5 per cent of normal, while UDPG-glycogen synthetase was approximately $^1/_2$ of normal. The phosphoglucomutase activity of a homogenate of the patient's muscle was about $^1/_3$ that of controls. The lower UDPG-glycogen synthetase and phosphoglucomutase are interpreted in terms of general degeneration of the tissue, while the lack of phosphorylase is considered to be a specific defect. The phosphorylase kinase activity of the homogenate was normal. The enzymatic pattern and the presence of a high concentration of glycogen in the tissue give strong evidence that glycogen is synthesized by way of UDPG.

The authors would like to acknowledge the capable technical assistance of Miss Lydia Hammaker.

* This work was supported in part by research grants from the National Cancer Institute and the National Institute of Arthritis and Metabolic Diseases, National Institutes of Health, U. S. Public Health Service, and the National Science Foundation.

The following abbreviations have been used: UDPG, uridine diphosphate glucose; AMP, adenosine monophosphate; TPN, triphosphopyridine nucleotide; TPNH, reduced triphosphopyridine nucleotide; Mg–ATP and Mn–ATP, the magnesium and manganous salts respectively of adenosine triphosphate; EDTA, ethylene diamine tetraacetic acid; Tris, tris(hydroxymethyl)–aminomethane hydrochloride.

[1] Sutherland, Jr., E. W., in *Enzymes: Units of Biological Structure and Function*, ed. O. H. Gaebler (New York: Academic Press, 1956), p. 536.

[2] Villar-Palasi, C., and J. Larner, *Biochim. et Biophys. Acta*, **30**, 449 (1958).

[3] Robbins, P. W., and F. Lipmann in CIBA Foundation Symposium on the "Regulation of Cell Metabolism" (London: J. A. Churchill, Ltd., 1959), p. 188.

[4] Robbins, P. W., R. R. Traut, and F. Lipmann, these PROCEEDINGS, **45**, 6 (1959).

[5] Leloir, L. F., and C. E. Cardini, *J. Am. Chem. Soc.*, **79**, 6340 (1957).

[6] Schmid, R., and R. Mahler, *J. Clin. Invest.*, **38**, 1040 (1959).

[7] Najjar, V. A., in S. P. Colowick and N. O. Kaplan, *Methods in Enzymology*, (New York: Academic Press, Inc., 1955), vol. 1, p. 294.

[8] Fischer, E. H., and E. G. Krebs, *J. Biol. Chem.*, **231**, 65 (1958).

[9] Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall, *J. Biol. Chem.*, **193**, 265 (1951).

[10] Illingworth, B., and G. T. Cori, in *Biochemical Preparations* (New York: John Wiley and Sons, Inc., 1953), vol. 3, 1.

[11] Wu, R., and E. Racker, personal communication.

[12] Krebs, E. G., and E. H. Fischer, *Biochim. et Biophys. Acta*, **20**, 150 (1956).

[13] Leloir, L. F., J. M. Olavarria, S. H. Goldemberg, and H. Carminatti, *Arch. Biochem. Biophys.*, **81**, 508 (1959).

[14] Dreyfus, J., G. Schapira, F. Schapira, and J. Demos, *Clinica Chimica Acta*, **1**, 434 (1956).

[15] Traut, R. R., unpublished experiments.

[16] Schmid, R., and R. Mahler, *J. Clin. Invest.* (in press).

[17] Dreyfus, J., G. Schapira, and F. Schapira, *J. Clin. Invest.*, **33**, 794 (1954).

[18] Nirenberg, M. W., *Biochim. et Biophys. Acta*, **30**, 203 (1958).

---

*HYDROXYMETHYLDEOXYCYTIDYLATE KINASE FORMATION AFTER BACTERIOPHAGE INFECTION OF ESCHERICHIA COLI**

BY RONALD SOMERVILLE,† KANEY EBISUZAKI, AND G. ROBERT GREENBERG

DEPARTMENT OF BIOLOGICAL CHEMISTRY, THE UNIVERSITY OF MICHIGAN

Communicated by Thomas Francis, Jr., June 8, 1959

Flaks and Cohen[1] have demonstrated that a new enzyme, deoxycytidylate hydroxymethylase, is formed after infection of *Escherichia coli* B by T-even bacteriophage. This finding opens the way for further investigation of the possible relationship of the DNA[2] of infecting phage to the synthesis of specific proteins and of the control of the formation of viral progeny.

We have observed the formation of another enzyme not detectable in normal cells. Hydroxymethyldeoxycytidylate kinase, which brings about the conversion of hydroxymethyldeoxycytidylate to the pyrophosphate level, is formed in a manner analogous to the synthesis of dCMP hydroxymethylase. At the same time an almost precipitous fall occurs in the ability of extracts to bring about net phosphorylation of dCMP. These enzymic changes can be considered as factors controlling the nature of the phage DNA.

Material and Methods.—Bacteriophage T2r was propagated on *E. coli* R2 in a glycerol-salts medium[3] and concentrated and purified by differential centrifugation.[4]

To prepare normal extracts *E. coli* R2 was grown aerobically on nutrient broth to 2–5 $\times$ 10^8 cells/ml, chilled, and harvested. The following manipulations were carried out at 0–4°. The washed cell paste was ground with several volumes of alumina, extracted with 0.05 *M* potassium phosphate buffer at pH 6.8, centrifuged at 25,000 $\times$ g and dialyzed overnight. In the experiments reported in Figure 1, the extracts were dialyzed against 0.01 *M* Tris buffer at pH 7.2, whereas in those experiments described in Tables 1 and 2 0.01 *M* potassium phosphate at pH 6.8 was used. For preparation of extracts of infected cells, *E. coli* R2 was grown with rapid circular shaking to 2–5 $\times$ 10^8 cells/ml in nutrient broth. The cells were infected at a multiplicity of about 5 with T2r and after 10 to 12 minutes

of continued shaking at 37°, were cooled, centrifuged, and ground and extracted as described above. In careful time studies 1-liter aliquots from a large shaking culture of infected cells were poured at intervals after infection into flasks containing ice at −20°. P^{32}-labeled *E. coli* DNA was prepared by the method of Lehman *et al.*[5] The DNA was degraded to 5′-monophosphates by a modification of the procedure of Lehman *et al.*[5] The deproteinized phosphodiesterase digest was separated into its component nucleotides on a Dowex-2-formate column by an ammonium formate gradient at pH 4.4, essentially by the procedure of Canellakis and Mantsavinos.[6] This procedure separated the 5 nucleotides cleanly. The deoxycytidylic acid was freed of ammonium formate by purification on a Dowex-50-H^+ column. Tetrahydrofolic acid was prepared by catalytic reduction of folic acid in glacial acetic acid, followed by lyophilization. It was dissolved in 0.05 *M* $KHCO_3$ and 0.02 *M* 2-mercaptoethanol under petroleum ether just prior to use.

HdCMP labeled in the $-CH_2OH$ group was prepared by incubation of infected extract with C^{14}-formaldehyde, tetrahydrofolate and dCMP essentially by the method of Flaks and Cohen.[1] P^{32}-labeled HdCMP was isolated from an enzymatic digest of P^{32}-labeled T2 DNA by chromatography on Dowex-2-formate columns. The C^{14}-HdCMP or P^{32}-labeled HdCMP was purified by chromatography on Dowex-50-H^+ ion exchange columns. Elution with water removes thymidylic acid in 3 bed volumes. HdCMP is separated cleanly in 16 bed volumes followed by dCMP at 34 bed volumes. Folic acid derivatives remain on the column, apparently still in combination with formaldehyde. This procedure using 0.01 *M* formic acid instead of water is the basis of a method for separating the C^{14}-labeled HdCMP to measure dCMP hydroxymethylase activity.[7] The HdCMP was characterized by its chromatographic behavior and by comparing its ultraviolet spectra at different pH values with those reported by Cohen[8] for hydroxymethyldeoxycytidine.

Deoxycytidylate kinase activity was measured according to the procedure of Friedkin and Kornberg.[9] Controls were prepared by heating 10 minutes in a 70 to 80° bath. In some instances HdCMP kinase activity was determined by using labeled HdCMP, chromatographing the reaction mixture on a Dowex-2-formate column and measuring the radioactivity of the isolated fractions. HdCMP, HdCDP and HdCTP were eluted respectively with 0.1 *M*, 0.45 *M*, and 0.6 *M* ammonium formate at pH 4.4. In the experiments described in Figure 1, HdCMP kinase activity was measured by passing the reaction filtrate through a Dowex-50-H^+ column and eluting with water. The di- and triphosphate derivatives passed directly through the column and their radioactivity was determined by plating aliquots. Under these conditions HdCMP is retained on the column as described above. The infected and normal extracts referred to in Table 2 were prepared from the same culture, a part being infected with bacteriophage at a multiplicity of 5 to 1. Experiments 1 and 2, Table 2, were carried out on successive days with the same extracts.

Results.—A kinetic study of the changes in these enzymes after infection is shown in Figure 1. In a separate experiment infection was effected at a multiplicity of 50 to 1 and extracts prepared at varying times after infection. Each extract was carried through a streptomycin treatment and a negative calcium phosphate gel adsorption followed by dialysis against phosphate buffer according to

the procedure of Lehman *et al.*[5] The same general trend was evident: that is, an increase in hydroxymethylase activity and a decrease in the apparent dCMP kinase activity with time.

Table 1 shows the results of a balance study of the new kinase reaction in a soluble extract of *E. coli* R2. HdCMP labeled with C^{14} in the $-CH_2OH$ group was the substrate in one experiment and $HdCMP^{32}$ the substrate in the other. In contrast to the extracts of the normal cells, which showed negligible phosphorylation of HdCMP, extracts of infected cells showed a considerable formation of HdCDP and rather less of HdCTP. The unaccountable loss of C^{14} perhaps suggests that the hydroxymethylase may be reversible and that the radioactivity was lost as a volatile non-charcoal absorbable compound; it is not due to self-absorption losses. The balance also showed that HdCMP was converted to an acid-insoluble product. Lehman *et al.*[5] have shown that such a conversion is due to a synthesis of DNA. This observation and the negligible incorporation of radioactivity into the acid-insoluble fraction in normal extracts is consistent with the absence of HdCMP kinase since HdCTP, once formed, would enter the DNA fraction[10] in competition with dCTP.

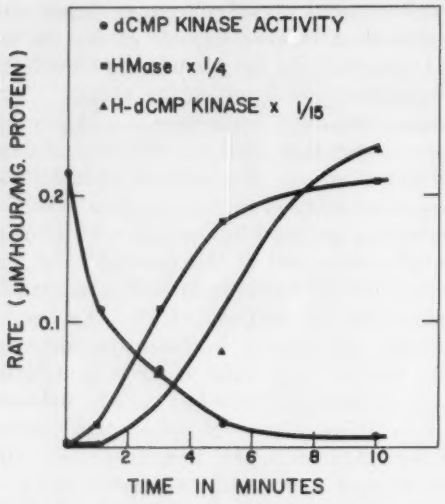

FIG. 1.—The dCMP kinase assay was essentially as described under Table 2. The reaction mixture for the assay of HdCMP kinase is described under Table 1 except that it contained 3 μmoles of ATP, 0.125 μmoles of C^{14}-HdCMP, 0.35 mg of protein from the zero time extract and decreasing quantities of protein as required for linear assay down to 5μg at 10 minutes. The reaction volume was 0.28 ml, reaction time 20 minutes. The temperature was 37°. The reaction system to assay dCMP hydroxymethylase contained the following: 10 μmoles of potassium phosphate, pH 7.6; 2.1 μmoles of dCMP; 4 μmoles of $MgCl_2$; 5 μmoles of 2-mercaptoethanol; 0.5 μmoles of formaldehyde containing 58,100 counts/minute/μmole; 0.11 ml. of a tetrahydrofolate solution (Table 2). The reaction was run under petroleum ether. The tetrahydrofolic acid and formaldehyde were preincubated 5 minutes before adding the enzyme. The reaction volume was 0.69 ml, reaction time, 40 minutes and temperature, 37°. For assay, 0.28 mg of protein from the zero time extract was used and then lesser quantities in the case of the infected extracts down to 0.07 mg of protein for the 10 minute determination. While it is not clear from the graph, the one minute value for HdCMP kinase had increased to 0.017 μmole/hr/mg.

Preliminary identification of the equilibrium mixture which was formed as a result of the HdCMP kinase activity was made by comparing its behavior to that of HdCMP, CDP, and CTP by paper electrophoresis, by Dowex-2-anionic column chromatography, and by paper chromatography.

Loss of ability to bring about net phosphorylation of dCMP is shown in Table 2, experiment 1. In the presence of formaldehyde, tetrahydrofolic acid, and ATP, infected extract catalyzed a rapid synthesis of nucleotide pyrophosphates from $dCMP^{32}$ (Experiment 2). This is evidence that HdCMP kinase activity was present in this extract.

TABLE 1

INCORPORATION OF HYDROXYMETHYL-LABELED HdCMP OR P^{32}-LABELED HdCMP INTO DI- AND TRIPHOSPHATE FRACTIONS AND ACID-INSOLUBLE FRACTION BY EXTRACTS OF INFECTED CELLS

		Total Counts per Fraction		
Fraction	Label	Normal	Infected	Infected Control
HdCMP	C^{14}	15,500	4720	12,320
	P^{32}	17,200	9260	15,880
HdCDP	C^{14}	0	2470	0
	P^{32}	221	5650	0
HdCTP	C^{14}	0	510	0
	P^{32}	*	474	*
Acid ppt.	C^{14}	3	314	4
	P^{32}	27	541	5
Charcoal non-adsorb.	C^{14}	72(66)	396(426)	42(66)
	P^{32}	432(420)	630(630)	192(142)

* A few scattered counts. The figures are counts/minute/fraction. Figures in parentheses represent counts after a second addition of charcoal. The reaction system contained the following: 2 μmoles of ATP; 5 μmoles of $MgCl_2$; 10 μmoles glycylglycine, pH 8.0; either infected or normal extract containing 3.25 mg of protein and 1.3 μmoles of potassium phosphate pH 6.8; and either 0.028 μmoles of C^{14} HdCMP or 0.029 μmoles of $HdCMP^{32}$. Total volume of the reaction mixture was 0.6 ml. temperature was 37°, and the reaction time 20 minutes. The reaction was stopped by adding 0.4 ml. of 20 per cent trichloroacetic acid. Trichloroacetic acid was added to the infected extract at time zero as a control.

TABLE 2

EFFECT OF EXTRACT OF INFECTED CELLS ON THE APPARENT dCMP KINASE ACTIVITY OF NORMAL EXTRACTS

Experiment	Extract	Ml	Treatment	P^{32} in Nucleotide Pyrophosphate Fraction. CPM
1	Normal	.01	None	657
	Normal	.02	None	1260
	Normal	.05	None	2400
	Normal	.05	Omit ATP	256
	Infected	.01	None	372
	Infected	.02	None	268
	Infected	.05	None	375
	Infected	.05	Omit ATP	205
2	Normal	.02	None	1295
	Normal	.02	Heated	41
	Infected	.05	Tetrahydrofolate $+CH_2O$	1079

The reaction system contained 10 μmoles of glycylglycine buffer, pH 8.0; 3 μmoles of ATP; 5 μmoles of $MgCl_2$; extract as shown and 0.34 μmoles of $dCMP^{32}$ with a specific activity of 40,000 counts/minute/μmole. The reaction volume was 0.26 ml, the reaction time, 20 minutes at 37°. The reaction was stopped by adding 0.5 ml. water and heating 2 minutes in a boiling bath. In addition to the above reactants, vessel 3 of experiment 2 contained 1.84 μmoles of formaldehyde and 0.05 ml of a solution containing 3 mg of tetrahydrofolic acid in 0.05 *M* $KHCO_3$ and 0.02 *M* 2-mercaptoethanol. Total volume was 0.36 ml. The normal and infected extracts contained 25 mg of protein/ml.

Discussion.—Following infection by bacteriophage a series of metabolic changes occur. It has been known for some time that DNase activity increases greatly.[11] The early work of Cohen[12] suggested that new protein synthesis took place in the first few minutes after infection and before DNA synthesis occurred. Experiments of Burton[13] with mutants admirably demonstrated that new protein synthesis is a necessary prelude to DNA synthesis. The demonstration by Flaks and Cohen[1] of the formation after infection of deoxycytidylate hydroxymethylase clearly established that at least one of the new proteins formed in this early period is a specific enzyme apparently not present in normal cells.

Studies in this laboratory have confirmed in every way the finding that dCMP hydroxymethylase[1] is formed after infection. In addition there is synthesized a specific kinase to convert HdCMP to the 5′-polyphosphate derivative. Furthermore net synthesis of dCTP or dCDP is rapidly abolished in infected cells (Fig. 1 and Table 2).

Recently Zimmerman *et al.*[14] and Koerner and Smith[15] have reported that after infection with T2 phage another new enzyme, dCTPase, is formed. This enzyme rapidly converts dCTP to CdMP and inorganic pyrophosphate. In addition, Zimmerman and co-workers[14] have found the HdCMP kinase reported here. This could explain the lack of net formation of deoxycytidine pyrophosphate derivatives in our experiments.

These changes in the enzymatic organization of the infected cells provide a remarkable control system to ensure that bacteriophage DNA will contain HdCMP and not dCMP. This becomes essential since it is apparent that DNA polymerase in the presence of phage DNA primer could not distinguish between dCTP and HdCTP.[10] Since dCMP hydroxymethylase is formed very early after infection any dCMP which might be formed from the ribonucleotide level[16] would rapidly be converted to HdCMP, thus removing it from the metabolic pool. However, an additional controlling device arises in the loss of ability to obtain net synthesis of 5′-polyphosphates from dCMP very shortly after infection.

While it is tempting to believe that these newly formed enzymes are initiated as a genetic function of DNA rather than as a result of a type of enzyme induction mechanism, no clear cut experiments are available to prove this. It is clear that whole bacteriophage formation is a genetic function and that DNA is involved.[17] Barner and Cohen[18] showed that a thymineless mutant, 15_T-, after infection with a T-even phage was able to synthesize net amounts of thymine as phage DNA. This finding is most readily interpreted to support the concept that the early protein synthesis—at least the synthesis of the missing enzyme in question—is dependent on a genetic site on the phage DNA. The observation that ultraviolet irradiation of phage destroys the ability of the phage to reproduce long before dCMP hydroxymethylase activity is lost[19, 20] may be explained by assuming that a specific array of genetic units is required for phage synthesis.[21]

It is pertinent to the discussion that with our present methods of assay we are unable to detect HdCMP kinase in normal extracts. While the method of assaying dCMP hydroxymethylase will detect less than 0.1 mμ moles of HdCMP synthesis per reaction vessel, no activity is detectable in normal extracts.[7] These findings tend to rule against an induction mechanism.[22]

Summary.—*Escherichia coli* R2 infected by bacteriophage T2r has been examined to determine whether enzymes other than deoxycytidylate hydroxymethylase are formed.

Experiments with hydroxymethyldeoxycytidylate labeled either with P^{32} or C^{14} have shown that this compound is converted to its di- and triphosphate derivatives in infected extracts. It is also converted to an acid-insoluble form, presumably DNA, in infected extracts.

The kinetics of formation of both deoxycytidylate hydroxymethylase and the new enzyme, hydroxymethyldeoxycytidylate kinase, have been described. These two enzymes are not detectable in normal extracts. Net conversion of deoxycytidylate to the 5′-polyphosphate derivative appears to be lost very rapidly after infection. The possible nature of the initiation of the synthesis of new enzymes following the infection of bacteria by T2r was discussed briefly.

We are deeply indebted to Dr. John Spizizen of the Department of Microbiology,

Western Reserve University, for supplying cultures of *E. coli* R2 and samples of bacteriophage T2r, for introducing us to the techniques of handling bacteriophage, and for helpful discussions.

* These studies have been supported by a research grant (A1669C) from the National Institutes of Health of the U. S. Public Health Service.

Some of the material in this paper is from a thesis to be presented by R. Somerville to the Horace H. Rackham School of Graduate Studies, The University of Michigan, in partial fulfillment of Ph.D. degree requirements.

† Special Scholar of the National Research Council of Canada.

[1] Flaks, J. G., and S. S. Cohen, *Biochim. Biophys. Acta*, **25**, 667 (1957).

[2] Abbreviations employed are: HdCMP, 5-hydroxymethyldeoxycytidine-5′-phosphate; HdCDP and HdCTP are the 5′-di- and triphosphate derivatives respectively; dCMP, deoxycytidine-5′-phosphate; dCDP and dCTP are the 5′-di- and triphosphate derivatives respectively; ADP and ATP, adenosine-5-diphosphate and triphosphate respectively; DNA, deoxyribonucleic acid; Tris, tris(hydroxymethyl)aminomethane.

[3] Frazer, D., and E. A. Jerrel, *J. Biol. Chem.*, **205**, 291 (1953).

[4] Spizizen, J., these PROCEEDINGS, **43**, 694 (1957).

[5] Lehman, I. R., M. J. Bessman, E. S. Simms, and A. Kornberg, *J. Biol. Chem.*, **233**, 163 (1958).

[6] Canellakis, E. S., and R. Mantsavinos, *Biochim. Biophys. Acta*, **27**, 643 (1958).

[7] Unpublished results. Ebisuzaki, K., S. Schlesinger, E. A. Figueiredo, and G. R. Greenberg, This method is approximately 100 times more sensitive than the original assay procedure.

[8] Cohen, S. S., *Cold Spring Harbor Symposium on Quantitative Biology*, **18**, 221 (1953).

[9] Friedkin, M., and A. Kornberg in *The Chemical Basis of Heredity*, ed. W. D. McElroy and Bentley Glass (Baltimore: The Johns Hopkins Press, 1957), p. 609.

[10] Bessman, M. J., I. R. Lehman, J. Adler, S. B. Zimmerman, E. S. Simms, and A. Kornberg, these PROCEEDINGS, **44**, 633 (1958).

[11] Pardee, A. B., and I. Williams, *Ann. Inst. Pasteur*, **84**, 147 (1953).

[12] Cohen, S. S., *J. Biol. Chem.*, **174**, 281 (1948).

[13] Burton, K., *Biochem. J.*, **61**, 473 (1955).

[14] Zimmerman, S. B., S. R. Kornberg, J. Josse, and A. Kornberg, *Federation Proc.*, **18**, 359 (1959).

[15] Koerner, J. F., and M. S. Smith, *Federation Proc.*, **18**, 254 (1959).

[16] Grossman, L., *Federation Proc.*, **17**, 235 (1958).

[17] Hershey, A. D., and E. Burgi, *Cold Spring Harbor Symposia on Quantitative Biology*, **21**, 91 (1956).

[18] Barner, H. D., and S. S. Cohen, *J. Bact.*, **68**, 80 (1954).

[19] Flaks, J. G., and S. S. Cohen, *Federation Proc.*, **17**, 220 (1958).

[20] Figueiredo, E. A., and S. Schlesinger, unpublished results.

[21] Doermann, A. H., M. Chase, and F. W. Stahl, *J. Cell. Comp. Physiol.* **45** (Suppl. 2), 51 (1955).

[22] A portion of the work reported here was presented before the Annual Meeting of The American Society of Biological Chemists: Somerville, R., and G. R. Greenberg, *Federation Proc.*, **18**, 329 (1959).

THE ROLE OF ORGANIC SUBSTRATES IN BACTERIAL PHOTOSYNTHESIS*

R. Y. Stanier, M. Doudoroff, R. Kunisawa, and R. Contopoulou

DEPARTMENT OF BACTERIOLOGY AND MILLER FOUNDATION, UNIVERSITY OF CALIFORNIA, BERKELEY

Communicated by C. B. van Niel, June 25, 1959

The essential character of bacterial photosynthesis was first clearly shown in 1930 by the investigations of van Niel on the green and purple sulfur bacteria.[1] Under anaerobic conditions in the light, these organisms reduce CO_2 to cell material with a concomitant stoichiometric oxidation of reduced sulfur compounds, as exemplified by the following equation:

$$CO_2 + 2H_2S \xrightarrow{\text{light}} (CH_2O) + H_2O + 2S.$$

In attempting to clarify the relations of such photosynthetic reactions to the reaction of green plant photosynthesis, van Niel[2] developed the theory that the central photochemical event in all photosyntheses is a cleavage of water to oxidized and reduced moieties, conventionally designated as [H] and [OH]. The reduced moiety serves as a reductant for conversion of CO_2 to cell material. In green plants, the oxidized moiety gives rise to molecular oxygen, and in bacteria it serves as the ultimate oxidant for the external substrate (Fig. 1). This simple and elegant

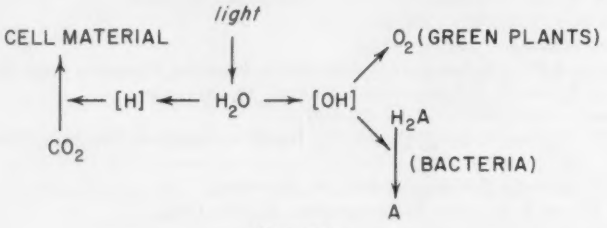

FIGURE 1.

scheme, proposed in 1941, has become generally accepted as an expression of the essential events in photosynthesis.

Many purple sulfur bacteria can grow anaerobically in the light in the absence of a reduced inorganic sulfur compound if they are provided with an organic substrate, and investigations of such photosynthetic processes were initiated shortly after van Niel's original work on the group.[3] Analyses of the chemical changes which occurred in cultures of purple sulfur bacteria furnished with organic substrates showed that the substrate is consumed with a simultaneous formation of cell material; there is also a slight net uptake or output of CO_2, the sign and magnitude of which depend on the oxidation level of the substrate. The general hypothesis of van Niel could be easily extended to cover this type of photosynthetic reaction (Fig. 2). The organic substrate is oxidized by photosynthetically generated [OH] with the formation of CO_2, and at the same time CO_2 is converted to cell material by reduction with photosynthetically generated [H]. The net CO_2 formation or up-

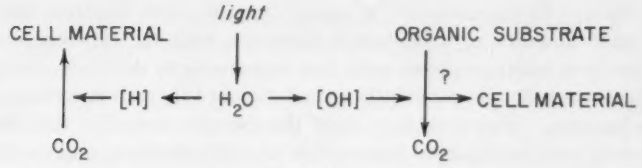

FIGURE 2.

take accordingly reflects the difference between two gross changes of opposite sign. It could not be excluded, of course, and was indeed explicitly suggested by van Niel,[3] that part of the organic substrate might under these circumstances be directly assimilated and converted to cell material; but the general theory required that at least part of it should serve as an ultimate photoreductant for the assimilation of CO_2, in order to provide a function for the photochemical cleavage of water with the generation of an oxidizing and a reducing agent which was assumed to be the unique photochemical feature of all photosyntheses.

In the meantime, Gaffron[4, 5] had initiated the biochemical study of a second group of purple bacteria, the non-sulfur purple bacteria. These organisms normally use organic substrates for photosynthesis, and unlike the purple sulfur bacteria, have complex growth factor requirements. At the time of Gaffron's work they could not be grown in media of known composition. Gaffron circumvented that difficulty by using manometric methods to measure photosynthetic transformations by resting cells. He showed that the photosynthetic utilization of fatty acids could be followed by measurements of CO_2 uptake by cells suspended in bicarbonate buffer. The observed uptake of CO_2 is caused by two different reactions: chemical CO_2 absorption resulting from disappearance of the organic anion; and biological CO_2 fixation. The latter must be determined indirectly, and the errors of measurement are considerable. Nevertheless, from repeated experiments with each substrate, Gaffron was able to infer the stoichiometry of these photosyntheses, and concluded that the assimilatory product formed by photosynthesis with fatty acids had the gross composition $C_4H_6O_2$. The best manometric evidence for this was obtained from experiments with acetate which did not involve a measurement of CO_2 changes. Gaffron discovered that acetate could be photometabolized with an uptake of molecular hydrogen in the absence of CO_2, the data corresponding to the equation:

$$2 \text{ acetate} + H_2 \xrightarrow{\text{light}} (C_4H_6O_2) + 2H_2O.$$

As formulated by Gaffron, the acetate-hydrogen reaction suggests a function for the organic substrate quite different from that postulated by van Niel. It can be interpreted as a *reductive photoassimilation* of acetate, completely divorced from CO_2 fixation. Gaffron in fact assumed that organic substrates are directly assimilated by purple bacteria in the light, and that any concomitant CO_2 fixation that occurs is determined by the amount of reducing power made available as a result of the assimilation of the organic substrate. In the case of acetate, which is more oxidized than the assimilatory product, no reduction of CO_2 should take place. Gaffron's hypothesis constituted an extension and refinement of a much older hypothesis first put forward without evidence by Molisch,[6] who had proposed that purple bacteria use light for the assimilation of organic compounds.

In an attempt to characterize the assimilatory product, Gaffron succeeded in isolating small amounts of a polymeric substance with the expected elementary composition from butyrate-grown cells, but was unable to determine its structure. He also discovered the photosynthetic assimilation of CO_2 with molecular hydrogen by purple bacteria. The stoichiometry of this reaction indicated that the assimilatory product had the empirical composition of a carbohydrate, and Gaffron therefore postulated that purple bacteria are capable of forming two assimilatory products with the empirical formulas $C_4H_6O_2$ and CH_2O. This assumption provided a means of explaining certain quantitative irregularities in CO_2 fixation at the expense of organic substrates, which are particularly conspicuous in Gaffron's data for the normal and branched isomers of fatty acids. Since all the isomers of a given fatty acid are on the same oxidation level, their photosynthetic conversion to a single assimilatory product should be accompanied by identical uptakes of CO_2. Instead, Gaffron found that the quantity of CO_2 assimilated per mole of fatty acid is systematically greater for branched isomers than for the normal isomer. Variations of this nature are understandable if the bacteria can form two assimilatory products on different oxidation levels, and if the nature of the assimilatory product is determined by the chemical structure of the organic substrate.

Gaffron's interpretation of the role of organic substrates in bacterial photosynthesis was a plausible one; but it failed to gain general acceptance, and was even tacitly abandoned by Gaffron himself some years later.[7] The concept of a direct photoassimilation of organic substrates did not fit very comfortably into any general theory of the nature of photosynthesis, as this process was understood between 1930 and 1950; in particular, it did not seem to permit a mechanistic formulation of a common photochemical event in all photosyntheses. Moreover, as van Niel very persuasively demonstrated[2, 8, 9] most of Gaffron's observations could be otherwise interpreted, by assuming that the organic substrate served as a photoreductant for CO_2. Lastly, a special case of bacterial photosynthesis with an organic substrate which functioned purely as a photoreductant was discovered in 1940 by Foster.[10] He succeeded in isolating an organism which performed an incomplete photosynthetic oxidation of isopropanol, a reaction which corresponded to the equation:

$$CO_2 + 2\begin{matrix}CH_3\\|\\CHOH\\|\\CH_3\end{matrix} \xrightarrow{\text{light}} (CH_2O) + H_2O + 2\begin{matrix}CH_3\\|\\C{=}O\\|\\CH_3\end{matrix}$$

This photosynthetic reaction provided a direct demonstration of the role assigned to organic substrates in photosynthesis by van Niel. However, it did not and could not prove that organic substrates that are completely consumed in the course of photosynthesis play a similar metabolic role. Further exploration of the photometabolism of such compounds had to await the introduction of carbon isotopes.

Experiments of Glover and Kamen[11] and Ormerod[12] on the photometabolism of acetate in the presence of $C^{14}O_2$ by resting cells of *Rhodospirillum rubrum* revealed that the endogenous CO_2 fixation is actually suppressed by the presence of

the organic substrate. This finding appeared all the more remarkable in view of the observation by Cutinelli *et al.*[13] that *Rhodospirillum rubrum* incorporates a substantial quantity of CO_2 in cellular proteins during growth in the presence of acetate and CO_2. Ormerod also observed that during the photometabolism of butyrate by resting cells of *R. rubrum*, the amount of $C^{14}O_2$ incorporated in the cells is identical with the fixation of CO_2 determined manometrically. Since butyrate is wholly consumed in the reaction, this fact shows that butyrate carbon must be directly and quantitatively incorporated into the cells. The photometabolism of these two fatty acids by resting cells therefore appears to conform rather well to the concept of a direct photoassimilation.

If, as the formulation of van Niel implied, organic substrates are in part completely oxidized to CO_2 during photometabolism, the enzymatic mechanisms of terminal respiration should continue to operate in purple bacteria furnished with organic substrates under anaerobic conditions in the light. This aspect of the problem was elegantly investigated by Elsden and Ormerod,[14] who examined the effects of fluoroacetate on the photosynthetic and respiratory metabolism of *R. rubrum*. Fluoroacetate proved to be an effective inhibitor of substrate oxidation in the dark, and its presence led to a substantial accumulation of citrate. The effects of fluoroacetate on the metabolism of the same organic substrate under anaerobic conditions in the light were in many cases strikingly different. Only with acetate, pyruvate, and oxalacetate did it markedly inhibit substrate utilization and cause an accumulation of citrate. Photometabolism of succinate, malate, and propionate was little affected. The rate of butyrate utilization was greatly reduced, but the inhibitor did not cause an accumulation of citrate. These findings strongly suggest, although they do not prove, that cyclic terminal respiration, universally operative in the dark oxidation of organic substrates by *R. rubrum*, ceases to be a major metabolic pathway in the light except in the case of acetate and closely related substrates. It can be further inferred that propionate, succinate, and malate undergo at most a very limited oxidation during photometabolism.

From this brief résumé, it will be evident that the role of organic substrates in bacterial photosynthesis is at present rather poorly understood. In this communication, we shall describe a series of experiments which permit a clearer insight into the problem.

Materials and Methods.—Rhodospirillum rubrum strain 1.1.1 from the collection of Professor C. B. van Niel was used in all experiments. Unless otherwise stated, cultures were grown in the basal medium of Hutner[15] under continuous aeration with 95% N_2–5% CO_2. The organic substrate in the standard Hutner medium is malate; in the present work, this was replaced by the specific substrate under study, which is indicated in the description of each experiment. Culture vessels were incubated at 30°C in an illuminated water bath at a light intensity of approximately 1000 f.-c.

For most of the experiments, it was important to use cells initially depleted as far as possible of endogenous carbon reserves. Such "depleted cells" were prepared in two ways; either by providing a growth-limiting concentration of the organic substrate, and harvesting the cultures 6–12 hours after the population had entered the stationary phase; or by harvesting cultures in the course of active growth, resuspending the cells in buffer, and incubating them under continuous aeration with

95% N_2–5% CO_2 for 6–12 hours in the light. The extent of depletion could be roughly ascertained by periodic cytochemical examination of the cells for the presence of sudanophilic granules and of polysaccharide.

In typical replenishment experiments, the depleted cells were resuspended at a density of approximately 1 mg (dry weight) per ml in phosphate buffer (0.02 *M*, pH 6.8), supplemented with the mineral base of Hutner.[15] At the start of the experiment, an initial sample of the suspension was withdrawn; the remainder of the suspension was furnished with an organic substrate, and incubated at 30°C in the light under continuous aeration with 95% He–5% CO_2. The cells were harvested after a substantial increase of the cellular dry weight had occurred. The period of replenishment varied in different experiments from 4 to 24 hours. The depleted and replenished cells were then analyzed in parallel. The cells were harvested by centrifugation, washed once with 0.005 *M* phosphate buffer (pH 6.8), and resuspended in a small volume of distilled water. Aliquots of this suspension were used for dry weight and radioactivity determinations and for chemical analyses.

The analytical procedure finally adopted was the following. The cell suspension was first heated at 100°C for 5 minutes to extract soluble materials. After centrifugation, the supernatant liquid was removed. This contained a small amount of carbohydrate, part of which was alcohol-precipitable polysaccharide. The residue was heated at 100°C for 60 minutes with 2 *N* H_2SO_4, in order to hydrolyze the water-insoluble polysaccharide of the cells. After centrifugation, the glucose content of the supernatant was determined by the use of glucose oxidase,[16] and the total sugar by the ferricyanide method.[17] In those experiments where lipid was determined, the insoluble residue remaining after acid hydrolysis was extracted successively with acetone and ether, and the dry weight of residue after evaporation of the combined extracts was measured. In most experiments, however, the residue was immediately digested for 12 hours at 37°C with alkaline hypochlorite[18] in order to isolate the granules containing poly-β-hydroxybutyric acid. These granules were then collected by centrifugation, washed successively with water, acetone-ether (1:2 v/v) and ether, and then extracted with hot chloroform. The poly-β-hydroxybutyric acid was determined gravimetrically on dried aliquots of the chloroform extract.

Direct isolation and purification of the reserve polysaccharide of *R. rubrum* were carried out by the procedures used for liver glycogen.[19]

Results.—Influence of the carbon source on the nature of the assimilatory products formed by purple bacteria: Some time ago, our colleague Dr. Germaine Cohen-Bazire observed that cells of *Rhodospirillum rubrum* and *Rhodopseudomonas spheroides* become packed with highly refractile sudanophilic granules during photosynthetic growth with acetate. Such granules cannot be seen in cells from comparable cultures grown with succinate or malate. If acetate-grown cells are broken by sonic oscillation, the granules can be isolated from the extract by differential centrifugation, sedimenting rapidly in the form of a chalky pellet which can be freed of other particulate cell-constituents by repeated washing on the centrifuge. Most of the material in the isolated granules is soluble in chloroform, and can be reprecipitated from the chloroform solution by the addition of a mixture of acetone and ether. This material has the characteristic melting point and infrared spectrum of poly-β-hydroxybutyric acid, a polyester long known as a cell-constituent of

certain *Bacillus* spp.,[20] and recently shown to occur in a wide variety of Gram negative non-photosynthetic bacteria.[21] The empirical formula for poly-β-hydroxybutyric acid, $(C_4H_6O_2)_n$, is identical with that assigned by Gaffron[5] to the unidentified assimilatory product formed during the photometabolism of fatty acids by purple bacteria. Furthermore, from the procedure which Gaffron used to isolate this product from cells of purple bacteria, certain of its properties can be inferred; and these fit well with the properties of poly-β-hydroxybutyric acid. Replenishment experiments with acetate-C^{14} and butyrate-C^{14}, described elsewhere,[22] show that much of the fatty acid carbon assimilated anaerobically in the light by *R. rubrum* is deposited in the cell as poly-β-hydroxybutyric acid. Hence there can be little doubt that the unidentified polymer found by Gaffron in purple bacteria was, in fact, poly-β-hydroxybutyric acid.

Cells of *R. rubrum* harvested during exponential photosynthetic growth with a series of organic substrates were analyzed for their polymer content (Table 1). In confirmation of earlier microscopic observations, these analyses show that polymer accumulation is markedly dependent on the nature of the organic substrate furnished. The virtual absence of polymer from cells grown with the C_4 dicarboxylic acids and with propionate led us to search for an additional assimilatory product.

TABLE 1

INFLUENCE OF THE ORGANIC SUBSTRATE ON THE CONTENT OF POLY-β-HYDROXYBUTYRIC ACID IN CELLS OF *R. RUBRUM*, HARVESTED DURING EXPONENTIAL GROWTH UNDER ANAEROBIC CONDITIONS AT 30°C IN THE LIGHT

Nature of Organic Substrate	Polymer Content of Cells, mg/100 mg Dry Weight
Acetate	23
DL-β-hydroxybutyrate	21
n-Butyrate	22
Pyruvate	3
DL-Malate	<0.1
Succinate	<0.1
Propionate	<0.1

TABLE 2

ASSIMILATORY PRODUCTS FORMED BY *R. RUBRUM* AS A RESULT OF PHOTOSYNTHETIC REPLENISHMENT WITH SUCCINATE FOR 16 HR IN HE-5% CO_2

	Weights, mg		
	Initial	Replenished	Change
Cells	600	750	+150
Cell constituents			
Ether-extractable lipid	43	39	−4
Poly-β-hydroxybutyric acid	3	10	+7
Water-extractable carbohydrates*	4	5	+1
Polysaccharide†			
Total reducing value	42	193	+151
Glucose	15	163	+148

* Reducing value by ferricyanide method on hot water extract of cells. Results expressed as glucose.
† Determined after hydrolysis of water-extracted residue with 2 *N* H_2SO_4 at 100°C for 1 hour. Total reducing value by ferricyanide method: glucose by enzymatic oxidation with glucose oxidase. Results expressed as glucose.

A culture of *R. rubrum* was grown photosynthetically with succinate, starved to deplete endogenous carbon reserves, and then used for a replenishment experiment with succinate. Comparative analyses of the starved and replenished cells (Table 2) showed that a small amount of poly-β-hydroxybutyrate had been formed; but this substance accounted for less than 5 per cent of the observed increase of dry weight. Examination of the cellular carbohydrates revealed that most of the in-

crease in dry weight was represented by the formation of polysaccharide that could not be extracted from the cells by hot water and was completely hydrolyzed by treatment of the cells at 100°C with 2 N H_2SO_4 for one hour. The native polysaccharide was evidently a polyglucose, since the increased sugar content of the acid hydrolysate from replenished cells could be wholly accounted for as glucose. The replenished cells stained intensely by the procedure of Hotchkiss[23] for the demonstration of polysaccharide, whereas the starved cells did not.

The polysaccharide was extracted from succinate-replenished cells and purified by the procedures employed for the isolation of liver glycogen.[19] Purified material yielded only glucose upon acid hydrolysis; on the basis of glucose content the best preparations were about 95 per cent pure. The polysaccharide stains a reddish brown with iodine, and is attacked by salivary and bacterial amylases. We conclude that it is a starch-like polysaccharide similar to glycogen and amylopectin.

Many subsequent replenishment experiments have been conducted with succinate, and polysaccharide has always been found as the major assimilatory product. A variable amount of poly-β-hydroxybutyric acid was also formed in replenishment experiments; it sometimes accounted for as much as 20 per cent of the increase in dry weight. The factors that control relative formation of the two products from succinate have not been investigated.

Gaffron[5] inferred that the assimilatory product formed as a result of photosynthesis with CO_2 and hydrogen had the empirical composition of a carbohydrate. We have performed one replenishment experiment with CO_2 and hydrogen; polysaccharide accounted for most of the assimilated carbon, but a substantial quantity of poly-β-hydroxybutyric acid was formed (Table 3). The replenishment period was very long, and it is possible that this may have caused an enhanced formation of poly-β-hydroxybutyrate. At all events, these experiments provide a general confirmation of Gaffron's observations and assumptions concerning the nature of the assimilatory products formed in bacterial photosynthesis.

TABLE 3

ASSIMILATORY PRODUCTS FORMED BY *R. RUBRUM* AS A RESULT OF PHOTOSYNTHETIC REPLENISHMENT WITH CO_2 AND H_2

	Dry Weights, mg			Per Cent of Total Dry Weight Increase
	Initial	After Replenishment	Increase	
Whole cells	1284	1536	252	
Cell constituents:				
Poly-β-hydroxybutyrate	5	86	81 } 252	32 } 100
Polysaccharide	195	366	171 }	68 }

The photometabolism of acetate by R. rubrum: Under replenishment conditions, poly-β-hydroxybutyric acid is the main assimilatory product formed during photometabolism of acetate, and acetate carbon enters the polymer virtually without dilution.[22] Accordingly, this biosynthesis can be grossly represented by the equation:

$$2n \text{ acetate} + 2n\ [H] \xrightarrow{\text{light}} (C_4H_6O_2)_n + 2_n\ H_2O.$$

The synthesis is reductive, and consequently some acetate must be oxidized in order to provide the necessary reducing power. This anaerobic oxidation of acetate probably proceeds through the tricarboxylic acid (TCA) cycle.[14] If all the available

hydrogen can be used, either directly or indirectly, for the reductive step in polymer synthesis, the balanced equation for the photosynthesis of poly-β-hydroxybutyrate from acetate is:

$$9n \text{ acetate} \xrightarrow{\text{light}} 4\ (C_4H_6O_2)_n + 2_n\ CO_2 + 2n\ H_2O.$$

Even under replenishment conditions, not all of the acetate assimilated appears as poly-β-hydroxybutyrate; the above equation is therefore an idealized one. There can be little doubt, however, that it represents the major reaction in resting cells. Ormerod's manometric data show a net photosynthetic CO_2 formation of 0.25 mole per mole of acetate metabolized,[12] in good agreement with the value of 0.22 mole required by the equation. Furthermore, the suppression of CO_2 fixation during the photometabolism of acetate by resting cells which was reported by Glover and Kamen[11] and by Ormerod[12] now becomes physiologically intelligible: the conversion of acetate to poly-β-hydroxybutyrate competes with CO_2 fixation for the limited reducing power available.

In principle, polymer synthesis from acetate should be possible without concomitant acetate oxidation, in the presence of an external reductant. Gaffron[5] observed a rapid uptake of hydrogen in the light by purple bacteria furnished with acetate in the absence of CO_2, a reaction which he formulated as a photoreduction of acetate:

$$2 \text{ acetate} + H_2 \xrightarrow{\text{light}} (C_4H_6O_2) + 2\ H_2O.$$

This interpretation was questioned by van Niel,[8] who pointed out that since hydrogen-adapted purple bacteria can reduce CO_2 very rapidly in the light, hydrogen uptake in the presence of organic substrates could be accounted for by the reaction:

$$CO_2 + 2\ H_2 \xrightarrow{\text{light}} (CH_2O) + H_2O,$$

occurring at the expense of carbon dioxide formed by oxidation of part of the organic substrate. The gross stoichiometry of acetate metabolism with hydrogen does not permit a clear-cut distinction between these two possibilities. We have accordingly reinvestigated this reaction.

A culture of *R. rubrum* was grown photosynthetically in an acetate medium, aerated with 95% H_2–5% CO_2. After exhaustion of the acetate, the aeration mixture was changed to 95% N_2–5% CO_2, and the culture was incubated for a further 20 hours in the light in order to deplete endogenous carbon reserves. The cells were harvested and replenished with 0.01 *M* acetate-C^{14}. Equal volumes of the resuspended cells were placed in three bottles, which were continuously aerated during replenishment with: pure hydrogen; pure helium; and helium containing 5% CO_2. The rates of acetate incorporation were determined periodically by removing aliquots of the suspensions and determining the radioactivity of the washed cells. The cells incubated with hydrogen were harvested after 4 hours, at which time they had incorporated 63 per cent of the acetate supplied. At this time, only about 15 per cent of the acetate had been incorporated by the cells incubated in helium and helium-CO_2, and the replenishment of these two suspensions was continued for a further 7 hours. When harvested after a total of 11 hours of replenishment, the cells incubated in helium had incorporated 22 per cent of the acetate

supplied, and the cells incubated in helium-CO_2 36 per cent. In each case, determination of the residual acetate in the supernatant liquid at the time of harvesting showed that most of the acetate which had been utilized had been incorporated in the cells. Consequently, the measurements of the rates of incorporation of radioactivity afford a valid measure of relative rates of acetate metabolism.

These results are incompatible with the assumption that hydrogen uptake reflects reduction of metabolically formed CO_2, which predicts that the rate of acetate utilization should be unaffected by the presence of hydrogen. The observed acceleration of acetate utilization by hydrogen is explicable in terms of a direct photoreduction, provided that acetate oxidation rather than polymer synthesis limits the over-all rate of acetate utilization.

In the cells which had been replenished in hydrogen, formation of poly-β-hydroxybutyrate accounted for 84 per cent of the increase in dry weight, and the isolated polymer contained 86 per cent of the total radioactivity of the cells. Analysis showed that a small increase in cellular carbohydrate had occurred, equivalent to 5 per cent of the dry weight increase; but the carbohydrate fraction, obtained by acid hydrolysis and freed of ionic material by Dowex treatment, was essentially devoid of radioactivity, containing at most 0.25 per cent of the total radioactivity of the cells. The cells replenished in helium showed a similarly high incorporation of acetate carbon in polymer, and a barely measurable incorporation in carbohydrate (about 1 per cent of the total in the cells). A markedly different incorporation pattern appeared in the cells replenished in helium-CO_2. Here the increase in carbohydrate accounted for over 20 per cent of the increase in dry weight, and the isolated carbohydrate fraction contained 9 per cent of the total C^{14} in the cells. Evidently, incorporation of acetate carbon into cellular carbohydrate requires the presence of CO_2, whereas incorporation into poly-β-hydroxybutyrate does not. This suggests that CO_2 may play an essential role in the formation of C_3 compounds from acetate by *R. rubrum*.

The fact that CO_2 is an important carbon source during photosynthetic growth of *R. rubrum* with acetate had already been established by Cutinelli *et al.*[13] They showed that 12 per cent of the carbon in the cellular proteins originates from CO_2 and 62 per cent from acetate when *R. rubrum* is grown in a medium containing acetate, bicarbonate, and yeast extract. Furthermore, the incorporation of CO_2 into the aminoacids is far from random; for example, the carboxyl group of alanine is derived almost exclusively from CO_2, while the α and β carbon atoms originate from the carboxyl and methyl groups of acetate, respectively. Such a distribution points to a formation of pyruvic acid by addition of CO_2 to an acetyl fragment.

The observations concerning the role of CO_2 fixation in biosynthesis which have just been described therefore suggested that CO_2 might be indispensable for the photosynthetic conversion of stored poly-β-hydroxybutyric acid to other cell materials. This was tested as follows.

Three aliquots of a suspension of acetate-replenished cells of *R. rubrum* in buffer were incubated anaerobically in the light. One was provided with a nitrogen source (0.1% NH_4Cl) and aerated with helium; one was provided with the same nitrogen source and aerated with helium containing 5% CO_2; and one was aerated with helium-5% CO_2 in the absence of a nitrogen source. After incubation for 16 hours, the suspensions were harvested and the changes in dry weight and cellular composi-

tion which had occurred were determined (Table 4). The changes in total dry weight were small and the cellular compositions expressed in percentile terms can therefore be validly compared. In the absence of CO_2, the polymer content of the cells fell slightly, but there was no increase in carbohydrate or nitrogen content, despite the presence of a nitrogen source. In the presence of CO_2 without a nitrogen source, the polymer content fell to about half its original value and the carbohydrate content doubled, but there was no change in nitrogen content. In the presence of both CO_2 and a nitrogen source, the polymer store was almost totally depleted, with concomitant increases in both carbohydrate and nitrogen content. The nitrogen content as presented in Table 4 is based on Kjeldahl nitrogen values; parallel determinations of protein by the Folin and the biuret methods also showed that protein increased in the presence of CO_2 and ammonia. This experiment accordingly demonstrates that CO_2 is essential for the photosynthetic conversion of stored poly-β-hydroxybutyric acid to carbohydrate and nitrogenous cellular constituents,

TABLE 4

THE INFLUENCE OF CARBON DIOXIDE AND AMMONIA ON THE PHOTOSYNTHETIC UTILIZATION OF ENDOGENOUS POLY-β-HYDROXYBUTYRIC ACID BY *R. RUBRUM*

		Replenished Cells, Before Starvation	After Starvation for 16 hr Anaerobically in the Light		
			With CO_2	With NH_3	With CO_2 and NH_3
Change in dry weight →			−2.5%	−5.5%	+1%
Cellular composition: milligrams per 100 mg cell material	(*a*) Poly-β-hydroxybutyrate	23	12	18	3
	(*b*) Carbohydrate*	9	18	10	18
	(*c*) Nitrogenous constituents†	50	49	48	59
	Sum of (*a*), (*b*), and (*c*)	82	79	76	80

* Measured as total reducing sugar on supernatant after hydrolysis of the cells for 60 min at 100°C with 2 *N* H_2SO_4.

† Determined by the Kjeldahl method, assuming an average nitrogen content of 16%.

including protein. We conclude that a photosynthesis of acetate will occur in the absence of CO_2; but the product of this reaction is a reserve material which can be used for general cellular synthesis only when CO_2 is simultaneously fixed. This explains the seemingly contradictory observations on CO_2 fixation by *R. rubrum* under resting and under growth conditions.

The photometabolism of butyrate: We have already reported[22] that photosynthetic replenishment of *R. rubrum* with butyrate-C^{14} results in the incorporation of much of the butyrate carbon in poly-β-hydroxbutyric acid, with little dilution of specific activity. The assimilation of butyrate can be formulated as:

$$n\ C_4H_8O_2 \longrightarrow (C_4H_6O_2)_n + 2n\ [H].$$

Manometric data[12] show that approximately 0.4 mole of CO_2 is reduced to cell material per mole of butyrate metabolized. Since Ormerod[12] observed identical net and gross fixation of CO_2 with butyrate, it follows that no butyrate carbon is evolved as CO_2. The CO_2 that is fixed photosynthetically with molecular hydrogen is converted principally to polysaccharide (Table 3). Accordingly, it may be expected that in the photometabolism of butyrate the two carbon sources, fatty acid and CO_2, are initially converted in the main to different assimilatory products, as shown by the following idealized equations:

$$2n\ C_4H_8O_2 \longrightarrow 2\ (C_4H_6O_2)_n + 4n\ [H]$$

$$4n\ [H] + n\ CO_2 \longrightarrow (CH_2O)_n + n\ H_2O$$

$$2n\ C_4H_8O_2 + n\ CO_2 \longrightarrow 2\ (C_4H_6O_2)_n + (CH_2O)n.$$

This was confirmed by conducting a replenishment experiment with C^{14}-labeled carbon sources. Two aliquots of a suspension of depleted cells grown with butyrate were furnished with 0.001 *M* butyrate and 0.001 *M* $NaHCO_3$, and incubated for 5 hours anaerobically in the light in completely filled, stoppered flasks. In one suspension butyrate was labeled with C^{14}, and in the other, bicarbonate. At the end of the replenishment period, the cells were analyzed and the fraction of the assimilated C^{14} which had been incorporated in the two major assimilatory products was determined (Table 5). As expected, butyrate carbon flows predominantly into polymer, and CO_2 carbon into polysaccharide.

TABLE 5

DERIVATION OF THE POLY-β-HYDROXYBUTYRATE AND POLYSACCHARIDE FORMED BY *R. RUBRUM* DURING PHOTOSYNTHETIC REPLENISHMENT WITH BUTYRATE AND CARBON DIOXIDE

Per cent of C^{14} Assimilated Found in	C^{14}-Labeled Substrate	
	Butyrate	Bicarbonate
(*a*) Poly-β-hydroxybutyrate	64	6
(*b*) Polysaccharide	10	66
(*c*) Sum of (*a*) and (*b*)	74	72

The photometabolism of succinate: Under replenishment conditions, the photometabolism of succinate results in synthesis principally of polysaccharide. In order to obtain some insight into the nature of this reaction, a replenishment experiment was conducted with C^{14}-labeled succinate. Two aliquots of succinate-grown cells were furnished respectively with succinate-1-C^{14} and succinate-2-C^{14}. Replenishment was continued until almost all the succinate had been metabolized, as determined by measurement of radioactivity on aliquots of the suspending medium, after removal of the cells by centrifugation. The replenished cells were fractionated to ascertain the general intracellular distribution of radioactivity. The polysaccharide, which was the most strongly labeled material in both batches of cells, was isolated directly and degraded by acid hydrolysis and subsequent alcoholic fermentation in order to determine the intramolecular distribution of C^{14}. The results of this experiment are shown in Table 6. Columns (*a*) and (*b*) give the derivations of cellular carbon from succinate-1-C^{14} and succinate -2-C^{14} respectively; column (*c*) gives the average derivation, calculated from the results in columns (*a*) and (*b*). From the average derivation, it can be calculated that 75 per cent of the assimilated succinate carbon enters polysaccharide, and 14 per cent enters poly-β-hydroxybutyric acid. The carbon from the carboxyl groups of succinate (column (*a*) is found predominantly in the two central carbon atoms of the hexose units of polysaccharide; very little enters poly-β-hydroxybutyric acid. The carbon from the methylene groups of succinate (column (*b*)) is found largely in the four outer carbon atoms of the hexose units of polysaccharide, but a substantial quantity also enters poly-β-hydroxybutyric acid. The total amount of carbohydrate extractable from the cells with hot water did not change appreciably as a result of replenishment. However, the sugar in the hot water extract was strongly labeled with C^{14}

TABLE 6

DISTRIBUTION OF SUCCINATE CARBON IN CELLS OF *R. RUBRUM* AS A RESULT OF PHOTOSYNTHETIC REPLENISHMENT WITH SUCCINATE-1-C^{14} AND -2-C^{14}

	Per Cent of C^{14} Metabolized		
	(*a*) Succinate-1-C^{14}	(*b*) Succinate-2-C^{14}	(*c*) $\frac{a+b}{2}$
Found in cells	34	77	56
In carbohydrate	30	53	42
In C_3 and C_4 of carbohydrate	28*	4	16
In C_1, C_2, C_5, C_6 of carbohydrate	2	49†	26
In polymer	1	14	8
In other cell materials (by difference)	3	10	6
Lost as CO_2 (by difference)	66	23	44

* Determined from specific activity of isolated glycogen and of CO_2 derived from it after hydrolysis and alcoholic fermentation.

‡ Determined from specific activity of isolated glycogen and of alcohol derived from it after hydrolysis and alcoholic fermentation.

TABLE 7

RELATIVE INCORPORATION OF CARBON ATOMS 1 AND 2 OF SUCCINATE IN POLYSACCHARIDE

	Specific Activities, cts/min/mg C.		Ratio of Specific Activities
	Succinate-1-C^{14}	Succinate-2-C^{14}	$\frac{2\text{-}C^{14}}{1\text{-}C^{14}}$
Substrate furnished	21,100	18,200	0.86
Glycogen isolated	8,050	14,500	1.80
Calculated for glycogen formed from succinate*	14,070	24,270	1.72

* Assuming loss of one carboxyl group, and conversion of resulting C_3 units to polysaccharide.

after replenishment, indicating that this pool is in partial equilibrium with the insoluble polysaccharide reserve.

The derivation of polysaccharide carbon suggests that hexose units are formed by the conversion of succinate to pyruvate, and subsequent hexose synthesis through a reversal of the glycolytic sequence. Some mixing occurs, so this cannot be the sole metabolic sequence that is operative; but it is certainly the predominant one. Were this pathway the sole mechanism for the conversion of succinate to polysaccharide, the ratio of the specific activities of glycogen carbon synthesized from succinate-1-C^{14} and succinate-2-C^{14} should be 1.72; the ratio actually found for the isolated glycogen was 1.80 (Table 7). The cells used for this experiment were not fully depleted, and the radioactivity of the polysaccharide was considerably diluted by the pre-existing non-radioactive pool. The exact correction which should be made for this pool is uncertain; but if a minimal value is assigned to it, calculation shows that the specific activity of the polysaccharide formed during replenishment is about 80 per cent of the specific activity calculated for polysaccharide synthesized from succinate alone. It therefore follows that at most 20 per cent of the polysaccharide was formed from CO_2; and the true magnitude of CO_2 incorporation may have been considerably less.

If acetyl units arise exclusively from succinate by oxidation via pyruvate, they should be derived only from the methylene carbon of succinate. Consequently, any poly-β-hydroxybutyric acid that is formed should originate from the central carbon atoms of succinic acid. Although most of the polymer synthesized does have this derivation, a small amount of carbon from succinate carboxyl groups also enters polymer. Furthermore, the specific activity of the polymer synthesized

from succinate-2-C^{14} is only 25 per cent of the expected value, assuming exclusive derivation from the two central carbon atoms of the molecule. No simple metabolic scheme can account for these results, and we conclude accordingly that a substantial part of the polymer formed during the photometabolism of succinate is not directly derived from the organic substrate.

In summary, this experiment indicates that during the photometabolism of succinate the substrate is initially oxidized to pyruvate. Most of the pyruvate is used for the synthesis of polysaccharide by reversal of glycolysis; a small amount is oxidized to acetate and used for polymer synthesis. The amount of substrate lost as CO_2 suggests that some succinate must have been completely oxidized; but the CO_2 production was calculated by difference, and the values shown in Table 6 are subject to considerable error. Since Elsden and Ormerod[14] have established that the photometabolism of succinate is grossly unaffected by fluoroacetate, the flow of substrate through the TCA cycle must be small.

Discussion.—The general inferences of Gaffron[4, 5] concerning the role of organic substrates in the photometabolism of purple bacteria have been confirmed and extended by the present study. It is now evident that the main function of organic substrates in bacterial photosynthesis is to serve as readily assimilable sources of carbon. They also provide reducing power for CO_2 fixation, as postulated by van Niel;[2] but the fixation of carbon dioxide, although qualitatively important for the growth of purple bacteria with organic substrates, is usually of minor *quantitative* importance as a source of cellular carbon.

When the rate of uptake of the exogenous carbon source is greater than the rate of gross cellular synthesis, the assimilated carbon is stored intracellularly by purple bacteria in the form of two principal reserve materials: poly-β-hydroxybutyric acid, a polyester more reduced than carbohydrate; and a glycogen-like polysaccharide. From any given exogenous carbon source, one of these reserve materials is formed in predominance. The chemical configuration of the substrate determines the nature of the assimilatory product. Substrates which are convertible to acetyl units without the intermediate formation of pyruvate, such as acetate and butyrate, yield mostly poly-β-hydroxybutyric acid. Substrates which can be directly converted to pyruvate with an accompanying generation of reducing power, such as succinate (and presumably also malate and propionate) yield mostly polysaccharide. The assimilation of carbon dioxide likewise results in polysaccharide formation.

The physiological function of polymer synthesis: In the absence of CO_2 practically all the acetate assimilated is converted to poly-β-hydroxybutyrate. If CO_2 is present during acetate assimilation, some acetate carbon is used for the synthesis of polysaccharide. Furthermore, CO_2 is essential for the utilization of stored poly-β-hydroxybutyrate: the polymer reserve cannot be drawn on for the synthesis of either carbohydrate or protein unless CO_2 is simultaneously furnished. These facts enable us to define the physiological function of polymer synthesis. It represents a means for the rapid and massive intracellular accumulation of exogenous organic carbon, when this organic carbon cannot be immediately converted to cell material. Free fatty acids obviously cannot be accumulated by the cell to any significant extent without causing serious damage. By conversion to polyester, these acids are neutralized and made osmotically inert, and the cell can build up a reserve of fatty acid carbon which represents a major fraction of its total dry weight.

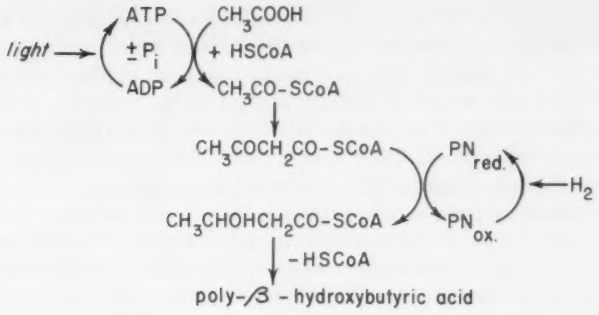

FIG. 3.

The polymer also constitutes an intracellular reserve of reducing power, which can be used for the subsequent fixation of CO_2. In this respect, its accumulation is functionally analogous to the intracellular storage of elementary sulfur that is characteristic of many purple bacteria.

The source of reducing power in photosynthesis with organic substrates: For many years, the photosynthetic oxidation of isopropanol discovered by Foster,[10] in which the organic substrate serves purely as a photoreductant of CO_2, was commonly used as a model to explain the role of the organic substrate in bacterial photosynthesis. This reaction can now be seen to represent an extreme and highly atypical case. It is instructive to analyze a contrasting reaction, the photometabolism of acetate, in which the organic substrate does not necessarily act at all as a photoreductant of CO_2. Acetate is converted anaerobically in the light by resting cells of *R. rubrum* largely to poly-β-hydroxybutyrate, this reductive synthesis being coupled either with an oxidation of part of the acetate, or with an uptake of molecular hydrogen. We shall now attempt an interpretation of the mechanisms of these photosynthetic reactions.

Although the enzymes which operate in the synthesis of poly-β-hydroxybutyrate from acetate by *R. rubrum* have not yet been investigated, it is a reasonable inference that the reductive step is the conversion of acetoacetyl-CoA to β-hydroxybutyryl-CoA, through electron transfer from reduced pyridine nucleotide.[24] In certain microbial systems, reduced pyridine nucleotide can be generated enzymatically as a consequence of the activation of molecular hydrogen.[25] It may therefore not be necessary to postulate for *R. rubrum* an electronic coupling through photochemically generated oxidant and reductant, as proposed in the scheme of van Niel.[2] Polymer synthesis from acetate requires a minimal input of one mole of ATP per mole of acetate, and if the electrons needed for the reductive step are transferred from hydrogen at the pyridine nucleotide level, no ATP can be formed as a result of the oxidation of hydrogen. Hence in this photosynthetic reaction, the only function which must be attributed to the light is the generation of ATP by photophosphorylation of the cyclic type.[26] This suggested interpretation of the acetate-hydrogen reaction is shown in Figure 3.

The situation when polymer synthesis is coupled with an oxidation of acetate can be interpreted in a similar fashion. If we assume that acetate is oxidized anaerobically through the TCA cycle, three pairs of electrons are removed at the pyridine

nucleotide level—in the oxidations of isocitrate, α-ketoglutarate and malate—and these oxidations can therefore be coupled *directly* with the reduction of acetoacetyl-CoA. The pair of electrons removed in the oxidation of succinate emerge at a higher potential, and cannot be directly applied to the reductive step in polymer synthesis. Since Frenkel[27] has shown that isolated chromatophores can catalyze a photochemical reduction of DPN, coupled with the oxidation either of succinate or of reduced flavin mononucleotide, it seems most probable that for this particular step the electronic coupling is linked through the primary photochemical reactions. Precisely as in the case of the acetate-hydrogen reaction, a massive generation of ATP by photo-phosphorylation is required for the synthetic reaction, since the cyclic oxidation of acetate cannot, under these circumstances, result in any net synthesis of ATP. From this analysis, we conclude that in bacterial photosynthesis the major function of light reactions is the generation of ATP. These reactions may be required for the generation of a photoreductant only when the oxidation of the external substrate yields electrons at a potential level higher than that of the pyridine nucleotides.

* This work was supported by grants from the National Science Foundation and the U. S. Public Health Service.

[1] van Niel, C. B., *Arch. Mikrobiol.*, **3,** 1 (1930).
[2] van Niel, C. B., *Advances in Enzymology*, **1,** 263 (1941).
[3] Muller, F. M., *Arch. Mikrobiol.*, **4,** 131 (1933).
[4] Gaffron, H., *Biochem. Z.*, **260,** 1 (1933).
[5] *Ibid.*, **275,** 301 (1935).
[6] Molisch, H., *Die Purpurbakterien nach neuen Untersuchungen*. (Jena: Gustav Fischer, 1907).
[7] Franck, J., and H. Gaffron, *Advances in Enzymology*, **1,** 199 (1941).
[8] van Niel, C. B., *Cold Spring Harbor Symposia*, **3,** 138 (1935).
[9] van Niel, C. B., *Arch. Mikrobiol.*, **7,** 323 (1936).
[10] Foster, J. W., *J. Gen. Physiol.*, **24,** 123 (1940).
[11] Glover, J., and M. D. Kamen, *Fed. Proc.*, **10,** 190 (1951).
[12] Ormerod, J. G., *Biochem. J.*, **64,** 373 (1956).
[13] Cutinelli, C., G. Ehrensvard, L. Reio, E. Saluste, and R. Stiernholm, *Ark. Kemi*, **3,** 315 (1951).
[14] Elsden, S. R., and J. G. Ormerod, *Biochem. J.*, **63,** 691 (1956).
[15] Cohen-Bazire, G., W. R. Sistrom, and R. Y. Stanier, *J. Cell. Comp. Physiol.*, **49,** 25 (1957).
[16] Measured by the use of Glucostat (Worthington Biochemical Corp.).
[17] Schales, O., and S. S. Schales, *Arch. Biochem.*, **8,** 285 (1945).
[18] Williamson, D. H., and J. F. Wilkinson, *J. Gen. Microbiol.*, **19,** 198 (1958).
[19] Somogyi, M., *Methods in Enzymology*, III (New York: Academic Press, 1957), pp. 3, 19.
[20] Lemoigne, M., *Ann. Inst. Pasteur*, **41,** 148 (1927).
[21] Forsyth, W. G., A. C. Hayward, and J. B. Roberts, *Nature*, **182,** 800 (1958).
[22] Doudoroff, M., and R. Y. Stanier, *Nature*, **183,** 1441 (1959).
[23] Hotchkiss, R., *Arch. Biochem.*, **16,** 131 (1948).
[24] Lynen, F., L. Wessely, O. Wieland, and L. Rueff, *Angew. Chem.*, **64,** 687 (1952).
[25] Korkes, S., *J. Biol. Chem.*, **216,** 737 (1955).
[26] Frenkel, A., *J. Am. Chem. Soc.*, **76,** 5568 (1954).
[27] Frenkel, A., *Brookhaven Symposia in Biology*, **11,** 276 (1959)

*SUBCELLULAR PARTICULATE SYSTEMS AND THE PHOTOCHEMICAL APPARATUS OF RHODOSPIRILLUM RUBRUM**

By Alice L. Tuttle and Howard Gest

THE HENRY SHAW SCHOOL OF BOTANY AND THE ADOLPHUS BUSCH III LABORATORY OF MOLECULAR BIOLOGY, WASHINGTON UNIVERSITY, ST. LOUIS

Communicated by F. W. Went, June 29, 1959

The photoactive pigments of the photosynthetic bacteria are believed to be localized in small spherical bodies called chromatophores.[1] Biochemical studies[2] have shown that chromatophore preparations catalyze a variety of light-dependent reactions (e.g., photophosphorylation of adenosine diphosphate) and they are, therefore, frequently considered to represent the functional equivalent of green plant chloroplasts. Electron micrographs[3] of thin sections of mature light-grown photosynthetic bacteria indicate that chromatophores are densely packed throughout the cell, and it has been known for some time that these bodies can be readily obtained from extracts of bacteria disrupted by sonic oscillation or mechanical abrasion. These observations have led to the impression that chromatophores are discrete morphological entities dispersed throughout the cytoplasm.

Sonic disruption of cells has been the most widely used method for preparing chromatophore suspensions. If sonication is extended beyond a critical time, however, photophosphorylation activity[4, 5] and the ability to produce molecular hydrogen photochemically[6] are lost. Continued sonic oscillation also leads to fragmentation of chromatophores.[7] In view of these results, it seemed likely that various properties of subcellular particles obtained by sonication might be significantly altered from those of the native photochemical system in the intact cell. Studies were therefore initiated to produce fragile cell-forms, such as protoplasts, of *Rhodospirillum rubrum*, from which the photochemical particulate system could be released by mild procedures.[8] These preliminary experiments disclosed that virtually all of the pigmented material in "lysates" (made by osmotic shock) of *Rhodospirillum* protoplasts was readily sedimented at low centrifugal force. Since free chromatophores obtained by the usual procedures are sedimented only at much higher forces (e.g., 46,000 to 54,000 $\times$ g for 50 minutes[2]) it was suggested that (*in vivo*) these bodies are associated with membranous components of the cell.

The present experiments with several types of altered cell-forms of *Rhodospirillum* offer further evidence that the system responsible for photochemical generation of adenosine triphosphate is normally associated with the cytoplasmic membrane or a system of membranous extensions penetrating the cytoplasm.

Materials and Methods. *Rhodospirillum rubrum* (strain S1) was grown photosynthetically at approximately 30°C in either the G3X or G5 media described by Kohlmiller and Gest.[9] Both media contain L-glutamic and DL-malic acids as major components, but differ with respect to their content of complex supplements (yeast extract, peptone).

Phase contrast examinations were made with the Leitz Ortholux microscope.

Rates of photophosphorylation of adenosine diphosphate (ADP) were estimated by determining utilization of inorganic phosphate as outlined by Frenkel.[4] Bacteriochlorophyll estimations were made using the methods of Cohen-Bazire *et al.*[10]

Protein concentrations (Fig. 1) were determined by the Folin-phenol method of Lowry *et al.*[11] on samples precipitated with trichloroacetic acid (crystalline bovine serum albumin used for standards).

Special reagents were purchased from the following sources: polymixin B sulfate, Nutritional Biochemicals Corp. (Cleveland, Ohio); phenazine methosulfate (PMS) and tris (hydroxymethyl) aminomethane (Tris), Sigma Chemical Co. (St Louis, Missouri); ADP, Sigma Chemical Co. and Pabst Laboratories (Milwaukee, Wisconsin); lysozyme, Sigma Chemical Co. and Armour Laboratories (Chicago, Illinois).

Results.—Protoplasts and "ghosts": Typical spiral-shaped cells of *R. rubrum* can be converted to nonmotile spherical protoplasts by procedures effective with a variety of other bacteria, e.g., by addition of penicillin (approx. 1000 units/ml) and high concentrations of an osmotic stabilizer such as sucrose to cultures actively growing in the light or by exposure of resting cells to lysozyme in the presence of versene (ethylenediaminetetraacetic acid) and sucrose. The lysozyme-versene method is technically more suitable for obtaining the quantities of *R. rubrum* protoplasts required for metabolic experiments and is described in detail in the preliminary communication of Karunairatnam, Spizizen, and Gest.[8] The latter investigators also compared several metabolic activities in intact cell and protoplast suspensions and observed qualitative similarity, except for the inability of protoplasts to photoevolve H_2. Absence of this activity may have been due to leakage of essential cofactors from the protoplasts, which very likely have an altered permeability as compared with whole cells.

R. rubrum lysozyme-protoplasts resemble those of other bacteria in many respects. Thus, dilution of concentrated suspensions (in 0.6 *M* sucrose) with water or addition of water to packed protoplasts usually yields gelatinous or stringy preparations. Considerable variability has been noted with different batches of protoplasts in this regard. In some instances, gelatinous preparations have been encountered during the final stages of protoplast preparation (with sucrose present). On the other hand, with other batches, the first treatment with water may yield "flaky" suspensions, and gels are sometimes observed only after several washes. These and numerous other observations suggest that *R. rubrum* protoplasts may not be as osmotically-fragile as the corresponding forms of Gram positive bacteria. The viscosity or "flakiness" observed in preparations made hypotonic by dilution with water is presumably due to release of desoxyribonucleic acid since it can be diminished or eliminated by addition of desoxyribonuclease.

Microscopic examination under phase contrast discloses that *R. rubrum* protoplasts enlarge upon dilution of sucrose suspensions with water and are converted to pale "ghosts" which usually appear to contain a number of opaque granules. These forms are strikingly similar in appearance to the "ghosts" derived from osmotically-shocked protoplasts of *Bacillus megaterium*.[12] Centrifugation of *R. rubrum* "ghost" suspensions at relatively low speeds (e.g., 14,500 × g for 20 minutes) results in a dark red-colored sediment and a pale pink supernatant liquid. Subsequent washing of the pellet with water frequently releases a small fraction of the pigmented material in a form not as readily sedimentable, but the bulk of the pigment remains associated with the "ghost" structures. The pigmented "ghosts" contain the functional photochemical apparatus which catalyzes photophosphoryla-

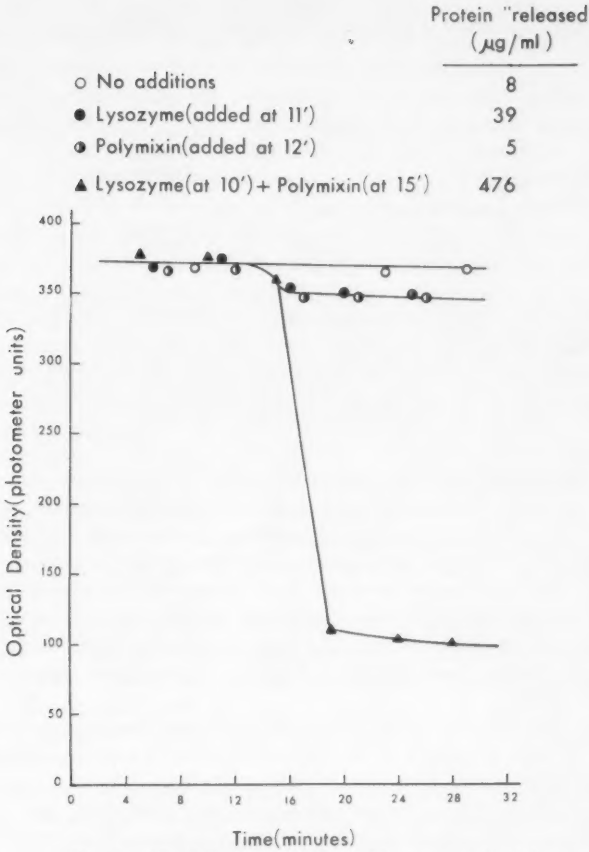

FIG. 1.—"Lytic" effect of sequential lysozyme-polymixin treatment on intact cells of *Rhodospirillum rubrum*. The cells (30 hours old) were grown photosynthetically in G5 medium, washed twice with 0.02 *M* Tris pH 8, and resuspended in 0.05 *M* Tris pH 8. 5.0-ml. aliquots (containing 9 mg dry weight of cells) of suspension were supplemented with 0.8 (▲), 0.9 (●, ◑), or 1.0 ml of water (○). Where indicated, 0.1-ml amounts of lysozyme (6 mg/ml) or polymixin B (2.5 mg/ml) solutions were added and the suspensions rapidly mixed. Optical densities were measured with the Klett-Summerson photoelectric colorimeter (No. 66 filter). Incubation temperature, 37°C. At 30 minutes, the suspensions were cooled, centrifuged to remove pigmented material, and the supernatant fluids analyzed for protein. The value given for protein "released" in curve ▲ may include some contribution, presumably 39 μg/ml or less, from unadsorbed lysozyme.

tion of adenosine diphosphate (Table 1).

The foregoing and similar experiments not detailed here indicate that photophosphorylation activity in "ghost" (or "lysed protoplast") suspensions shows many over-all similarities with the process as it occurs in preparations of free chromato-

TABLE 1

PHOTOPHOSPHORYLATION OF ADENOSINE DIPHOSPHATE BY "GHOSTS" OF *R. RUBRUM*

Vessel	"Cofactor" Additions*	Atmosphere	μmoles Inorganic Phosphate Esterified per Hour per mg Bacterio-chlorophyll
1	None	Helium	0
2	None	Hydrogen	0
3	Mg^{++} (20 μmoles) + succinate (0.5 μmole)	Helium	16.9
4	Mg^{++} + succinate	Hydrogen	15.5
5	Mg^{++} + succinate + phenazine methosulfate (25 μg/ml)	Helium	57.7

* *Incubation conditions:* Warburg vessels of approximately 10 ml capacity, containing a final fluid volume of 2 ml were employed. Each vessel contained 0.8 ml of "ghost" suspension, 400 μmoles glycylglycine pH 7.6, 15 μmoles inorganic phosphate, and 10 μmoles adenosine diphosphate. Other additions as listed above. Temperature, 30°C. Light intensity, ~1000 foot-candles.

"Ghost" preparation: Light-grown cells (40 hours old, from G3X medium) were converted to protoplasts by the ysozyme-versene method. 10 ml of protoplast suspension (in 0.6 *M* sucrose; derived from the cells in 280 ml of medium) were centrifuged at 14,500 × *g* for 20 min and the pellet resuspended in 10 ml. of water. 0.6 mg of Worthington desoxyribonuclease was added to the "flaky" suspension and after incubation at 30°C for 5 min, the suspension was recentrifuged. The resulting pellet was washed once more with water, resuspended in 10 ml of water, and uniformly dispersed using a hand-operated Ten Broeck homogenizer.

phores.[4, 5] Thus, with both types of preparations the reaction is light-dependent, requires the presence of an acceptor such as adenosine diphosphate, is stimulated by catalytic amounts of phenazine methosulfate (provided succinate is present at low concentration) and is not significantly influenced by the presence of a hydrogen atmosphere. Mg^{++} and small amounts of succinate, or other reducing agents, are required for optimal photophosphorylation by purified free chromatophores but not by crude preparations obtained by sonic oscillation.[4] From the results of Table 1, it appears that these or similar "cofactors" must be lost during preparation of the washed "ghosts."

Protoplasts and "ghosts" of *R. rubrum* are relatively fragile structures and can be disrupted in various ways with the apparent release of chromatophores, e.g., by shaking with Ballotini beads in the Mickle apparatus or by exposure to surface active agents such as sodium lauryl sulfate. Salton[13] has noted that isolated cell wall preparations of *R. rubrum* and *E. coli* appear to be dissolved by sodium lauryl sulfate and similar anionic detergents. The possible usefulness of sodium lauryl sulfate as a dispersing agent for *Rhodospirillum* altered cell-forms is, however, somewhat limited by the fact that the detergent interacts, more or less rapidly depending on its concentration, with the pigment system causing marked visible color changes (the 880 mμ absorption peak of bacteriochlorophyll may also be significantly depressed).

Effects of polymixin: The structure of the outer integument of typical Gram negative bacteria appears to be considerably more complex than the cell walls of Gram positive organisms such as the bacilli.[14] One of the outstanding differences between the two types is the presence of a high proportion of lipid or lipoprotein in the surface structures of Gram negative cells. The studies of B. A. Newton,[15] Few,[16] and others indicate that the cyclic polypeptide antibiotic polymixin (polymyxin) reacts with lipid components of cytoplasmic membranes of sensitive Gram negative bacteria, causing disorganization of the membrane and an attendant loss of specific permeability properties. It might be expected that polymixin would also affect lipoprotein wall components. These considerations prompted tests with this

antibiotic and it was found that exposure of resting cells of *Rhodospirillum*, in 0.05 *M* Tris buffer pH 8, to 250 μg/ml of polymixin B resulted in the formation of granulated protoplast-like bodies over the course of several hours.

Since polymixin can induce the formation of protoplast-like bodies and other aberrant morphological forms from spiral-shaped cells, it seems likely that lipid components of the wall mosaic must play a significant role in maintaining the structural rigidity of the cell. This conclusion is also supported by the facts that cells of *R. rubrum* growing in the presence of 60 μg polymixin B/ml (in G3X medium) are bacilli-form, rather than spiral-shaped, and that treatment of normal spirals with 10 per cent phenol destroys the characteristic shape yielding rod-like morphological elements.

"Circle" preparations: Protoplasts or protoplast-like bodies produced from *R. rubrum* are ordinarily quite opaque as viewed under phase contrast. In numerous types of preparations, e.g., of intact cells frozen and thawed in the presence of lysozyme and in "ghost" suspensions treated with sodium lauryl sulfate, we have observed spherical bodies which appear to be completely transparent and sharply delineated by a thin membrane; an adhering patch of phase-dense material is sometimes revealed by careful focusing. These forms closely resemble the empty "ghosts" which Weibull[17] prepared by treatment of the Gram positive *Bacillus megaterium* with lysozyme in the absence of an osmotic stabilizer. Because of their microscopic appearance we use the operational term "circle" to represent such structures.

"Circles" can be rapidly made in almost 100 per cent yield by briefly exposing spiral-shaped cells to lysozyme and then to polymixin B under the conditions given for Figure 1. With lysozyme alone, the spiral morphology is unaffected but a small number of protoplasts are invariably observed in the suspensions.[18] Polymixin alone under these conditions, does not cause observable morphological alterations.

When a suspension of normal spirals is converted to "circles" by the synergistic action of lysozyme and polymixin B, the optical density decreases rapidly upon addition of the antibiotic and is attended by an extensive release of protein (see Fig. 1). The suspension clarifies and superficially appears to have lysed completely. Control suspensions with no supplements, or with lysozyme or polymixin added individually, do not show significant decreases in optical density or liberation of protein.

The "lysis" of *R. rubrum* by sequential lysozyme-polymixin treatment is greatly retarded if the order of addition is reversed. This effect is reminiscent of Repaske's observations[19] that, although the addition of versene and Tris stimulated rapid lysis by lysozyme of *Escherichia coli*, *Azotobacter vinelandii*, and several pseudomonads, in the case of *E. coli* at least, preincubation with versene before lysozyme addition markedly reduced the rate of lysis. The interpretation was nevertheless offered that versene exerts its effect by chelating a metal on the cell surface, thereby exposing the "lysozyme-substrate." Warren et al.[20] similarly have suggested that polymixin causes exposure of "lysozyme-substrate," since in their studies on the combined action of lysozyme and polymixin on several Gram negative bacteria, increased susceptibility to lysis was achieved by pretreatment with polymixin. It would be of interest, in view of the present findings with *R.*

rubrum, to determine whether the organisms examined by Warren et al. are lysed even more rapidly by preexposure to lysozyme before polymixin addition. Perhaps these organisms would also be converted to "circles" by such treatment.

The experimental results with *R. rubrum* appear to lend themselves to the interpretation that the "lysozyme-substrate" in the cell surface of this organism is readily accessible, but is involved together with polymixin-sensitive components (presumably lipid or lipoprotein) in conferring the characteristic shape to the intact normal cell. It is conceivable that versene, like polymixin, interacts with lipid constituents of the cell "envelope," and that, in fact, prior disruption of these elements may somehow retard the action of lysozyme.

Practically all of the pigmented material in "circle" suspensions of *R. rubrum* is sedimented by centrifugation at 6,300 × g for 15 minutes. In contrast with intact cells and protoplasts, the pigmented "circle" pellet is translucent and is similar in this respect to pellets of free chromatophores. It may be noted that "circles" do not contain any of the large opaque granules characteristic of *R. rubrum*, which argues against the possibility of physical entrapment of free chromatophores by the limiting (damaged) membrane. "Circle" preparations catalyze the photophosphorylation of adenosine diphosphate as illustrated by the data of Table 2.

TABLE 2

PHOTOPHOSPHORYLATION OF ADENOSINE DIPHOSPHATE BY "CIRCLES" PREPARED FROM *R. RUBRUM*

Tube	Incubated in	"Cofactor" Additions*	Δ μmoles Inorganic Phosphate per 40 Minutes†
1	Light	None	0
2	"	Mg^{++} (20 μmoles)	−0.3
3	"	Succinate (0.5 μmole)	−0.4
4	"	Phenazine methosulfate (25 μg/ml)	−1.5
5	"	Mg^{++} + succinate	−0.4
6	"	Mg^{++} + succinate + PMS	−4.3
7	Dark	Mg^{++} + succinate + PMS	+0.2

* *Incubation conditions:* As in Table 1, except that the suspensions were incubated in square Thunberg tubes, open to the atmosphere. Each tube contained 0.5 ml of "circle" preparation, made as described below.

† Change from zero time control.

"Circle" preparation: Light-grown cells (24 hours old) derived from 500 ml of G5 medium were converted to "circles" under the conditions given for Figure 1 (▲), but at a greatly increased scale (750 ml starting volume). A considerable amount of floating, pigmented material was removed from the surface of the "lysed" suspension by careful siphoning and the remainder then centrifuged at 1,460 × g for 10 min to remove stringy pigmented masses. "Circles" were collected from the supernatant fluid by centrifugation at 7,000 × g for 15 min. The resulting pellet was washed once with 20 ml of 0.05 *M* glycylglycine buffer pH 7.6 (containing 1.9 μmoles inorganic phosphate per ml), resuspended in 7 ml of the same buffer, and homogenized with a hand-operated Ten Broeck.

With all three "cofactors" (Mg^{++}, succinate, and PMS) present, the rate of photophosphorylation shown by this particular preparation was 77 μmoles inorganic phosphate esterified per hour per mg bacteriochlorophyll. In the absence of the dye, however, "circle" preparations show unexpectedly low activity. Petrack[21] has recently reported that cell-free extracts of blue-green algae show a similar dependence on PMS for active photophosphorylation.

At the present time it is not known whether the photochemically active "circles" produced from *R. rubrum* by the combined action of lysozyme and polymixin represent cytoplasmic membranes, membranes with adhering surface components, or several layers of the integument which are considerably disorganized but still fused together. Analysis of a circle preparation for hexosamine,[22] a characteristic cell surface constituent, revealed its presence to the extent of approximately 80 per cent of the amount found in normal cells (~400 μg, calculated as glucosamine, per

100 mg dry weight of cells). This finding is of interest but, considering our inadequate knowledge of the structure of the Gram negative cell integument, does not permit an unambiguous choice among the several possibilities.

Treatment of typical "circle" suspensions with sodium lauryl sulfate causes a decrease in optical density (No. 66 filter), resulting from disintegration with the release of pigmented fragments. For example, in an experiment with 1 mg detergent/ml, a negligible pellet was obtained by centrifugation at 26,000 × g for 30 minutes; the bacteriochlorophyll in the supernatant fluid was markedly affected, however (ratio of optical density at 880 mμ to optical density at 800 mμ ≅ 1.3). With 0.5 mg/ml, a thin translucent pellet was observed and the pink supernatant fluid showed a corresponding optical density ratio of approximately 6, which is characteristic for bacteriochlorophyll in chromatophores of *R. rubrum*.[1] It is possible that under proper circumstances sodium lauryl sulfate, or similar surface active agents, may release active "sub-units" from the supporting membranous structures.

Discussion.—Many investigators have noted that an appreciable fraction of the photosynthetic pigments in biochemically active cell-free extracts of purple bacteria is sedimentable at relatively low centrifugal force. Ordinarily, this material is discarded and free chromatophores are isolated (by high speed centrifugation) from the supernatant fluids resulting from centrifugation at 25,000 × g for 30 minutes or longer. The present experiments demonstrate that mild lytic procedures, such as osmotic shock of protoplasts, do not release chromatophores in significant quantity, but rather that the functional photophosphorylation apparatus remains associated with large membranous structures which are sedimentable at low speeds. It is therefore suggested that the photoactive pigment system of the photosynthetic bacteria is associated with the (lipoprotein) cytoplasmic membrane and/or a reticulum of membranous extensions penetrating the cytoplasm. A corollary of this interpretation is that the bacterial pigment complex does not normally exist in the form of independent cytoplasmic inclusions comparable to the chloroplasts of higher plants.

The foregoing conception would reconcile our findings with a number of observations which until now seemed contradictory. Studies[3] of thin sections of light-grown photosynthetic bacteria by electron micrography indicate that chromatophores, in cells grown in the usual manner, are densely packed throughout the cytoplasm. On the other hand, the results of immunochemical (and other[7]) studies with *Chromatium* by J. W. Newton[23] have led him to conclude that "the photosynthetic apparatus of purple bacteria is not necessarily a unique, discrete intracellular entity, since it contains macromolecular configurations common to the cell surface on one hand and certain 'intracellular' proteins on the other. It seems advisable, therefore, provisionally to consider it a part of a more complex organizational state within the cell." It is pertinent to note that "cell wall" preparations obtained from *R. rubrum* by the usual technique (disruption in the Mickle shaker) show only slight pigmentation. The presence of a small amount of pigment could be explained as being due to contamination of the outer cell-envelope by adhering inner membrane structures. Few[24] has, in fact, reported that prolonged shaking of Gram negative bacteria in the Mickle disintegrator is required to produce cell walls devoid of the "secondary inner layer."

The mechanism and dynamics of chromatophore formation are still obscure. It is of special interest in this connection that Frenkel[25] has recently observed that *very young* light-grown cells of *R. rubrum* do not contain chromatophores, but instead contain smaller (200–300 Å) photochemically active particles. Furthermore, similar particles are obtained if older cells grown in the usual way are sonicated in viscous media.[2, 25] These pigmented particles are reported to be as photochemically active as chromatophores. It is tempting to speculate that the small "photosynthetic sub-units" are derived from the cytoplasmic or similar membranes. Cytoplasmic membranes of bacteria are fragile structures, known to be readily disintegrated by sonic oscillation and other vigorous disruption procedures[26, 27] and, accordingly, it has been suggested that various subcellular particulate species observed in bacterial extracts made by these techniques represent fragments of such membranes. This view is supported by increasing evidence that many "particle-bound" enzyme systems can also be shown to be integral parts of cytoplasmic membranes[26, 28] or the cell "envelope."[29]

Summary.—Osmotic lysis of altered fragile cell-forms (protoplasts) of the Gram negative photosynthetic bacterium *Rhodospirillum rubrum* does not release significant amounts of free chromatophores, but results in the formation of pigmented membranous structures ("ghosts") which are capable of photophosphorylating adenosine diphosphate. A second type of photochemically active form, which closely resembles the "empty ghosts" of certain Gram positive bacteria, can be prepared by the sequential action of lysozyme and the polypeptide antibiotic polymixin B on resting cells. Both types of preparations, in contrast with free chromatophores, are readily sedimented at low centrifugal force. On the basis of the current findings and other observations from the literature, it is concluded that *in vivo* the photochemical apparatus of the photosynthetic bacteria is associated with the cytoplasmic membrane and/or membranous extensions in the cytoplasm of the cell.

* This investigation was supported by a research grant (E-2640) from the National Institute of Allergy and Infectious Diseases, United States Public Health Service and was initiated while the authors were members of the Department of Microbiology, Western Reserve University School of Medicine, Cleveland, Ohio.

[1] Schachman, H. K., A. B. Pardee, and R. Y. Stanier, *Arch. Biochem. Biophys.*, **38**, 245 (1952).

[2] Frenkel, A. W., in *The Photochemical Apparatus, Its Structure and Function* (Brookhaven Symposia in Biology, No. 11 (1959)), p. 276.

[3] Vatter, A. E., and R. S. Wolfe, *J. Bacteriol.*, **75**, 480 (1958); Bergeron, J. A., in *The Photochemical Apparatus, Its Structure and Function* (Brookhaven Symposia in Biology, No. 11 (1959)), p. 118.

[4] Frenkel, A. W., *J. Biol. Chem.*, **222**, 823 (1956).

[5] Geller, D. M., Ph.D. thesis, Harvard University, Cambridge, Mass. (1957).

[6] Karunairatnam, M. C., and H. Gest, unpublished observations.

[7] Newton, J. W., and G. A. Newton, *Arch. Biochem. Biophys.*, **71**, 250 (1957).

[8] Karunairatnam, M. C., J. Spizizen, and H. Gest, *Biochim. et Biophys. Acta*, **29**, 649 (1958); in keeping with a practice suggested by A. G. Jose, Jr., and P. W. Wilson, these PROCEEDINGS, **45**, 692 (1959), we employ the term protoplast as such for the forms referred to, with the understood qualification that they are not entirely analogous to the protoplasts of Gram positive bacteria.

[9] Kohlmiller, E. F., Jr., and H. Gest, *J. Bacteriol.*, **61**, 269 (1951).

[10] Cohen-Bazire, G., W. R. Sistrom, and R. Y. Stanier, *J. Cell. Comp. Physiol.*, **49**, 25 (1957).

[11] Lowry, O. H., N. J. Rosebrough, A. L. Farr, and R. J. Randall, *J. Biol. Chem.*, **193**, 265 (1951).

[12] Weibull, C., *Exptl. Cell Research*, **10**, 214 (1956).

[13] Salton, M. R. J., in *Proceedings of the Second International Congress of Surface Activity* (New York: Academic Press, Inc., 1957), vol. 4, p. 245; see also Bolle, A., and E. Kellenberger, *Schweiz. Z. Path. Bakt.*, **21**, 714 (1958).

[14] Salton, M. R. J., *J. Gen. Microbiol.*, **18**, 481 (1958); Weidel, W., and J. Primosigh, *J. Gen. Microbiol.*, **18**, 513 (1958).

[15] Newton, B. A., in *The Strategy of Chemotherapy*, ed. S. T. Cowan and E. Rowatt (Cambridge: Cambridge University Press, 1958), p. 62.

[16] Few, A. V. in *Proceedings of the Second International Congress of Surface Activity* (New York: Academic Press, Inc., 1957), vol. 4, p.288.

[17] Weibull, C., *J. Bacteriol.*, **66**, 688 (1953).

[18] During phase contrast examination of cells (in 0.05 *M* Tris pH 8) treated with lysozyme alone we have also observed the apparent extrusion of "circles" from protoplasts and other events strikingly reminiscent of the plasmoptysis phenomenon described by Alfred Fischer in 1900 (*Zeit. für Hyg.*, **35**, 1 (1900)) in a remarkable study of the effects of osmotic disturbances on a variety of bacteria.

[19] Repaske, R., *Biochim. et Biophys. Acta*, **30**, 225 (1958).

[20] Warren, G. H., J. Gray, and J. A. Yurchenco, *J. Bacteriol.*, **74**, 788 (1957).

[21] Petrack, B., *Fed. Proc.*, **18**, 302 (1959).

[22] By an adaptation of the method described by Boas, N. F., *J. Biol. Chem.*, **204**, 553 (1953).

[23] Newton, J. W., in *The Photochemical Apparatus, Its Structure and Function* (Brookhaven Symposia in Biology, No. 11 (1959)), p. 289.

[24] Few, A. V., *J. Gen. Microbiol.*, **10**, 304 (1954).

[25] Frenkel, A. W., *Ann. Rev. Plant Physiol.*, **10**, 53 (1959).

[26] Weibull, C., *J. Bacteriol.*, **66**, 696 (1953).

[27] Newton, B. A., *J. Gen. Microbiol.*, **12**, 226 (1955); Mitchell, P., and J. Moyle, in *Bacterial Anatomy*, ed. E. T. C. Spooner and B. A. D. Stocker (Cambridge: Cambridge University Press, 1956), p. 150.

[28] Mitchell, P., and J. Moyle, *Biochem. J.*, **64**, 19P (1956); Storck, R., and J. T. Wachsman, *J. Bacteriol.*, **73**, 784 (1957).

[29] Marr, A. G., and E. H. Cota-Robles, *J. Bacteriol.*, **74**, 79 (1957).

---

*A CRITERION FOR ORBITAL HYBRIDIZATION AND CHARGE DISTRIBUTION IN CHEMICAL BONDS**

BY MARTIN KARPLUS AND D. M. GRANT†

NOYES CHEMICAL LABORATORY, UNIVERSITY OF ILLINOIS, URBANA

Communicated by Nelson J. Leonard, June 15, 1959

Although parameters related to orbital hybridization and charge distribution have long been used in descriptions of the chemical bond,[1] reliable values for these quantities have been very difficult to obtain. In this report, we suggest that electron-coupled nuclear spin interactions, when interpreted in terms of the theory of localized electron pairs, can serve to measure the extent of orbital hybridization and bond polarization.

The coupling constant $A_{NN'}$ (in units of cps) for a pair of nuclei with spins $\mathbf{I}_N$ and $\mathbf{I}_{N'}$ is defined by the expression

$$E_{NN'} = hA_{NN'}\,\mathbf{I}_N \cdot \mathbf{I}_{N'} \qquad (1)$$

where $E_{NN'}$ is the interaction energy, which can be determined by an analysis of the

field-independent multiplets observed in the high-resolution NMR spectra of molecules.[2, 3] With approximate second order perturbation theory,[4] the coupling constant $A_{NN'}$ can be written in a form which involves the appropriate Hamiltonian operator[5] and the ground state molecular wave function Ψ_0. As Ψ_0 for the $2n$ electron system, we choose a function of the form[6, 7]

$$\Psi_0 = [(2n)!]^{-1/2} \sum_P (-1)^P \{\psi_a(1, 2)\psi_b(3, 4) \ldots \psi_m(2n - 1, 2n)\} \qquad (2p$$

where P includes all permutations of the $2n$ electrons and the $\psi_r(i, j)$ are orthonormal products of the symmetric space functions $u_r(i, j)$ and antisymmetric spin functions; i.e.,

$$\psi_r(i, j) = [2]^{-1/2} u_r(i, j) \{\alpha(i)\beta(j) - \beta(i)\alpha(j)\} \qquad (3)$$

Substitution of the expression for Ψ_0 (Eqs. 2 and 3) and for the Hamiltonian of Ramsey and Purcell[5] into the general integral for $A_{NN'}$ yields the result

$$A_{NN'} = \frac{\gamma_N \gamma_{N'}}{3h} \Big\{ \sum_r < u_r(i, j) \mid O_1(i, i) + O_{1b}(i, j) + O_2(i, i) + O_2(i, j) + O_3(i, j) \mid u_r(i, j) > + \sum_{r, s > r} < u_r(i, j)u_s(k, l) \mid 2\, O_{1b}(i, k) + O_2(i, k) \mid u_r(i, j)u_s(k, l) > - \sum_{r, s > r} < u_r(i, j)u_s(k, l) \mid O_{1b}(i, k) + O_2(i, k) - O_3(i, k) \mid u_r(k, j)u_s(i, l) > \Big\} \qquad (4)$$

where

$$O_{1a}(i, i) = \frac{8e^2\hbar^2}{2mc^2} \frac{\mathbf{r}_{iN} \cdot \mathbf{r}_{iN'}}{r^3_{iN} r^3_{iN'}} \; ; \; O_{1b}(i, j) = -\frac{8\hbar^2}{\Delta_{1b}} \frac{\mathbf{m}_{iN} \cdot \mathbf{m}_{jN'}}{r^3_{iN} r^3_{jN'}}$$

$$O_2(i, j) = \frac{8(\beta\hbar)^2}{\Delta_2} \frac{3(\mathbf{r}_{iN} \cdot \mathbf{r}_{jN'})^2 - r^2_{iN} r^2_{jN'}}{r^5_{iN} r^5_{jN'}}$$

$$O_3(i, j) = \frac{1}{\Delta_3} \left(\frac{16\pi\beta\hbar}{3}\right)^2 \delta(\mathbf{r}_{iN}) \delta(\mathbf{r}_{jN'}) \qquad (5)$$

and $O_1 = O_{1a} + O_{1b}$. The symbols used in Equation (5) have their standard meanings with Δ_p representing the average excitation energy for the operator O_p.

The formulation given in Equations (4) and (5) can be applied to the calculation of molecular coupling constants for situations in which deviations from perfect pairing are not important.[8] This is generally the case for electron-spin interactions between directly bonded atoms.[9] For a detailed treatment of such a system, a choice for the form of $u_r(i, j)$ has to be made. We employ a generalized valence-bond function,

$$u_r(i, j) = \eta(\lambda_x, \lambda_y) \{\phi_x(i)\phi_y(j) + \phi_x(j)\phi_y(i) + \lambda_x\phi_x(i)\phi_x(j) + \lambda_y\phi_y(i)\phi_y(j)\} \qquad (6)$$

where ϕ_x and ϕ_y are hybrid orbitals on nucleus x and y, respectively, λ_x and λ_y are parameters determining the charge distribution and $\eta(\lambda_x, \lambda_y)$ is the appropriate normalizing factor with inclusion of overlap. With this form of $u_r(i, j)$, the cou-

pling constant $A_{NN'}$ becomes a function of the degree of hybridization in the atomic orbitals ϕ_x, ϕ_y and the bond polarizations λ_x, λ_y. Usually the calculation can be considerably simplified without significant loss of accuracy by including in Equation (4) only the $u_r(i, j)$ which have at least one of the atomic orbitals on nucleus N or N'. This approximation is made in the calculations reported below.

The simplest application of the theory is to systems of high symmetry in which the hybridization and/or bond polarities are known. Although there are few, if any, molecules in which both types of parameters have been determined, the structures of BH_4^-, CH_4, and NH_4^+ are such that the central atom in each system can be assumed to contribute tetrahedral hybrids to the localized X-H valence bonds ($X = B, C, N$). Further, only s and p orbitals are likely to make a significant contribution to the hybrids, which can be written in the form $\phi_x = \frac{1}{2} s_x + \frac{\sqrt{3}}{2} p_{\sigma x}$. For these molecules, the contribution of O_1 and O_2 terms to the coupling is quite small[10] and can be neglected in an approximate treatment. Also, in the $u_r(i, j)$ function used here, λ_H was set equal to zero[11] and λ_x ($X = B, C, N$) became the only unknown parameter. To evaluate the integrals in Equation (4), use was made of Hartree-Fock atomic orbitals, undistorted by bond formation.[12] The average excitation energy Δ was set equal to 9 e.v.; this value is approximately twice the X = H bond energy.[13, 14] Table 1 lists the calculated results for A_{HX} as a function of λ_x^2. The experimental values obtained by Lauterbur[15] are also given, as are the values of λ_x^2 obtained by fitting the calculated A_HX to the experimental results. It is evident from the Table that general agreement with experiment is obtained for reasonable λ_x^2 values and, further, that the relative magnitudes of λ_x^2 are in the order expected from electronegativity concepts. A quantitative comparison is difficult since there exists no exact relationship between electronegativities and bond polarization, as expressed here in terms of the λ_x with inclusion of overlap. Perhaps the quantity $(\eta^2\lambda_x^2)$ might be said to have the closest correspondence to the usual "ionic character"; values for $(\eta^2\lambda_x^2)$ are given in the last column of Table 1.

TABLE 1

THE COUPLING CONSTANT A_{HX} (CPS) FOR TETRAHEDRAL MOLECULES

Molecule	A_{HX} (calc.) $\lambda x^2 = 0$	$\lambda x^2 = .05$	$\lambda x^2 = .1$	$\lambda x^2 = .2$	$\lambda x^2 = .3$	A_{HX} (exp)*	λ_x^2 (exp)†	$[\eta^2\lambda_x^2]$‡
BH_4^-	105	84.7	77.3	67.2	...	80.5 ± .5	.07	.02
CH_4	181	147	134	117	...	124 ± 3	.14	.047
NH_4^+	89.3	72.9	66.6	57.9	51.2	53.5 ± .5	.27	.096

* The experimental values for A_{HX} are from the unpublished measurements of P. C. Lauterbur (private communication). For BH_4^-, there are some earlier results by Ogg, R. A., [*J. Chem. Phys.*, **22**, 1933 (1954)] which range from 77–87 cps.

† The values of λ_x^2 were obtained by fitting A_{HX} (calc.) to A_{HX} (exp.)

‡ η is the normalization constant defined in Equation (6).

As an additional application of the theory, we treat some non-tetrahedral systems whose NMR spectra have been analyzed and X-H coupling constants determined. For such molecules, an evaluation of the extent of hybridization is much more difficult than in the tetrahedral case. To avoid this complication, we assume that the hybrid atomic orbitals are directed along the bond line. If two or more equivalent X-H are present, the orbital hybridization can then be determined by the orthogonality requirement and the measured bond angles.[16] The three molecules

TABLE 2

COUPLING CONSTANT A_{HX} (CPS) FOR NON-TETRAHEDRAL MOLECULES

Molecule	HXH angle	a^2§	A_{HX} (calc)	A_{HX} (exp)	λ_x^2 (exp)
B_2H_6 (non-bridge H)	121.5°*	.337	109	125 ± 2‖	.025
C_2H_4	117.4°†	.315	156	157 ± 2¶	.14
NH_3	106.8°‡	.224	48	46 ± 2**	.30

* Hedberg, K., and V. Schomaker, *J. Am. Chem. Soc.*, **73**, 1482 (1951).
† Allen, H. C., Jr. and E. K. Plyler, *J. Am. Chem. Soc.*, **80**, 2673 (1958).
‡ Herzberg, G., *Infrared and Raman Spectra* (New York: D. Van Nostrand Company, Inc., 1954); Weiss, M. T., and M. W. P. Strandberg, *Phys. Rev.*, **83**, 567 (1951).
§ a^2 is the square of the s orbital coefficient ($\phi_x = as_x + bp\sigma_x$).
‖ Ogg, R. A., Jr., *J. Chem. Phys.*, **22**, 1933 (1954).
¶ Muller, N., private communication.
** Ogg, R. A., Jr., and J. D. Ray, *J. Chem. Phys.*, **26**, 1515 (1957).

B_2H_6 (non-bridge H), C_2H_4, and NH_3 with accurately known structures and coupling constants are chosen for discussion. In Table 2 the experimental bond angles and a^2 values ($\phi_x = as_x + bp^4_x$) determined from them are given. To simplify the treatment, we again neglect O_1 and O_2 contributions to the coupling and assume that the average excitation energy Δ is unaltered from the value used for the XH_4 molecules. If the λ_x^2 values found for the XH_4 systems are used for the molecules under discussion, the theoretical A_{HX} values listed in Table 2 are obtained. Comparison with the experimental coupling constants shows good agreement for the C—H bond[17] but some difference for the *B—H* and *N—H* bonds. Since for the latter two systems one is comparing a charged with a neutral species, a change in λ_x^2 might be expected. Adjustment of the λ_x^2 values to obtain agreement between A_{HX} (calc.) and A_{HX} (exp.) gives the results listed in the last column of Table 2. The differences for the λ_x^2 values between BH_4^- and B_2H_6 and between NH_4^+ and NH_3 are so small that their significance is doubtful.[18] However, the general agreement is such as to indicate that the hybrid orbitals do indeed point at least approximately along the bond line. The import of this result is somewhat tempered by the drastic approximations employed in the theory. An additional fact that should be mentioned is that although O_1 terms are certainly small for the tetrahedral case, their magnitude increases with deviations from tetrahedral symmetry. More refined calculations are required to determine the contribution of these terms in other systems.

In conclusion, it can be stated that the results obtained with the localized electron pair theory indicate that the coupling constant $A_{NN'}$ is a useful criterion for the study of bond properties, since it is sensitive to the amount of hybridization and covalency in the wave function.[19] As in other methods (e.g., nuclear quadrupole coupling[20]), the interpretation is complicated by the presence of parameters that are difficult to evaluate. It would be of great interest, therefore, to examine molecules of such a nature that a number of criteria can be used simultaneously. Measurements on compounds containing boron or other nuclei for which both the nuclear quadrupole coupling and the spin-spin coupling can be determined would be of particular interest in this regard. Results on such systems will be given in future reports.

We wish to thank P. C. Lauterbur and N. Muller[21] for providing us with the results of their measurements prior to publication. We also wish to thank Dr. J. M. Parks for his cogent comments on the manuscript.

* This research was supported by a grant from the Petroleum Research Fund administered by the American Chemical Society. Grateful acknowledgment is hereby made to donors of said fund. Also, the work was supported in part by the Office of Naval Research and by a grant from the University of Illinois Research Board.

† Now at the Department of Chemistry, University of Utah, Salt Lake City, Utah.

[1] Pauling, L., *The Nature of the Chemical Bond* (Ithaca: Cornell University Press, 1939).

[2] Gutowsky, McGall, and Slichter, *Phys. Rev.*, **84,** 589 (1951).

[3] Hahn, E. L., and D. E. Maxwell, *Phys. Rev.*, **84,** 1246 (1951).

[4] Lennard-Jones, J. E., *Proc. Roy. Soc.* (London) *A***129,** 598 (1930).

[5] Ramsey, N. F., and E. M. Purcell, *Phys. Rev.*, **85,** 143 (1952); Ramsey, N. F., *Phys. Rev.* **91,** 303 (1953).

[6] Hurley, A. C., J. E. Lennard-Jones, and J. A. Pople, *Proc. Roy. Soc.* (London) *A***220,** 446 (1953).

[7] Parks, J. M., and R. G. Parr, *J. Chem. Phys.* **28,** 335 (1958).

[8] For a formulation that is valid if deviations from perfect pairing are significant, see Karplus, M., and D. H. Anderson, *J. Chem. Phys.* **30,** 6 (1959).

[9] A qualitative approach, which foreshadows the ideas expressed more quantitatively in this paper, has been given by Gutowsky, McCall, and Slichter (see Ref. 2).

[10] In CH_4, we have estimated $O_1 \approx 2$ cps and $O_2 \approx 8$ cps. It should be pointed out that the calculation of these two numbers required more effort than all of the other results put together. Details will be reported elsewhere.

[11] Coulson, C. A., *Trans. Farad. Soc.*, **38,** 433 (1942).

[12] The use of the free atom functions is valid for B, C, and N but may introduce an error for hydrogen. Since orbital contraction is likely in ϕ_H, the calculated values of A_{HX}, are probably somewhat low.

[13] Karplus, M., *J. Chem. Phys.*, **30,** 11 (1959).

[14] Cottrell, T. L., *The Strengths of Chemical Bonds* (London: Butterworth's Scientific Publications, 1954).

[15] Lauterbur, P. C., private communication.

[16] Coulson, C. A., *Valence* (Oxford: Oxford University Press, 1952).

[17] For acetylene, the use of the ideal (sp) hybrid with $a^2 = 1/2$ yields a coupling constant of 250 cps in agreement with the values (247–251 cps) obtained by Lauterbur for some substituted acetylenes.

[18] The changes required in λ_x^2 for the cases considered are consistent with each other; they are, however, in a direction *opposite* to that which might be expected on the basis of a simple picture. A small change in the orbital charge also could account for the observed results.

[19] The dependence of the coupling constant on the s-orbital contribution to the wave function can be determined by comparing the A_{HX}(exp) value in Table 1 with the A_{HX} (calc) value in Table 2.

[20] Townes, C. H., and B. P. Dailey, *J. Chem. Phys.*, **17,** 782 (1949).

[21] N. Muller has kindly informed us that he has found an empirical correlation between s–orbital hybridization and the coupling constant in CH bonds.

THE EFFECTS OF SPREADING TECHNIQUE AND PURITY OF SAMPLE ON THE EVAPORATION RESISTANCE OF MONOLAYERS

By Victor K. La Mer and Geoffrey T. Barnes

DEPARTMENT OF CHEMISTRY, COLUMBIA UNIVERSITY

Communicated June 15, 1959

Until recently, the objective of studies on the rate of evaporation of water through surface films has been primarily a search for adequate experimental techniques. The best apparatus was designed by Langmuir and Schaefer[1] who placed a circular absorbing box a few millimeters above the water surface of a film balance. The bottom of the box was a water permeable membrane which supported a desiccant. The box was removable and thus could be weighed at intervals giving the rate of evaporation. With this form of apparatus the film surface pressure can be varied at will, and. more important still, the resistances to the passage of water vapor arising from the measuring system are evaluated by direct measurement. Thus, for the first time, the specific evaporation resistance of the monolayer alone was determined; this resistance is independent of all apparatus constants and is expressed in absolute units.

Unfortunately the results of Langmuir and Schaefer were not reproducible and often exhibited hysteresis on expansion and compression of the film. Archer and La Mer,[2] continuing this work with the same type of apparatus, suggested that impurities in the monolayer were responsible for the erratic results of Langmuir and Schaefer. They introduced several improvements in the method, particularly in the technique of spreading the monolayer, in order to prevent contamination by the atmosphere, the underlying subphase, and the spreading solvent. Their efforts were successful and reproducible results were then obtained.

To prevent impurities from entering the monolayer between the final cleaning of the water surface and the first measurements, Archer and La Mer spread the monolayer at a high initial surface pressure. They claimed that the conventional method of spreading at zero surface pressure permitted the entry of impurities and thus would not give reproducible results. To minimize the risk of retaining some of the solvent molecules in the monolayer (a recognized source of error in their early experiments) the spreading solutions were made more concentrated than those normally used (4 mg/ml as compared to 1 mg/ml).

Rosano and La Mer[3] verified the results of Archer and La Mer when they used this special spreading technique; but with some modifications (p. 349).

However, if reliable results could be obtained by conventional spreading methods, it would be possible to extend the studies of evaporation resistance with more certainty to mixed monolayers. The Archer-La Mer technique is unsuitable for spreading mixed monolayers because crystals are inevitably formed in the monolayer during spreading. Thus, if one of the components crystallized preferentially the composition of the remaining monolayer would not then be known.

Experimental.—The apparatus for measuring evaporation resistance was the same as that described by Archer and La Mer,[2] except that surface pressures were measured with a Wilhelmy dipping plate of roughened mica suspended from a rod on a torsion wire.[3] The monolayers were spread using an "Agla" micrometer syringe (Burroughs Wellcome & Co. London).

The precautions we have taken to prevent contamination of the monolayers are listed below:

(*a*) the water used in the Langmuir trough was redistilled in a dust-proof apparatus using a column packed with glass rings;

(*b*) the spreading solvent, petroleum ether (40–50°C fraction), was twice distilled;

(*c*) the trough and other glass-ware were cleaned with concentrated nitric acid (*not* with sulfuric-chromic cleaning mixture);

(*d*) only the barriers and the edges of the trough were lightly waxed following the technique of Archer and La Mer[2] (later the glass barriers were replaced by Teflon strips);

(*e*) the surface balance was enclosed in a wooden box with hinged plexiglass covers which were coated lightly on the inside with vaseline to retain dust particles and prevent them from falling into the trough;

(*f*) the entire apparatus was enclosed in a vaselined hood;

(*g*) the room was air conditioned, with dust filters;

(*h*) several monolayers were spread on the water surface and swept off before evaporation measurements were commenced, according to the recommendations of Langmuir and Schaefer.[1]

Using these precautions, monolayers of stearic acid were spread under three different conditions and the rates of evaporation measured. They involved spreading:

(1) from a solution of high concentration (4 mg/ml) at a high initial surface pressure (>10 dynes/cm), as used by Archer and La Mer[2, 3];

(2) from a solution of high concentration (4 mg/ml) at zero initial pressure;

(3) from a low concentration (1 mg/ml) at zero initial pressure (the conventional method).

The results are listed in Table 1. If the monolayers spread under conditions (1) and (2) were viewed in oblique light at a proper angle small crystals could be seen scattered throughout the *entire* monolayer and not just at the point of spreading as reported by Archer and La Mer.[2] No crystals could be observed in monolayers spread by method (3).

In a second series of experiments stearic acid and octadecanol monolayers were each spread by the conventional method (3) and the evaporation resistances measured as functions of surface pressure. In these experiments only three resistance measurements were made on any one monolayer and the sequence of pressures was randomized. In Figures 1 and 2 the results are compared with those obtained by Archer and La Mer and by Rosano and La Mer using method (1).

Calculation of Evaporation Resistance.—From the measured rates of evaporation (m/t) it is possible to calculate specific evaporation resistances (r) in two ways.

(*a*) Using the equation of Archer and La Mer,[2]

$$r = [a(w_w - w_d)t/m]_f - [a(w_w - w_d)t/m]_w \qquad (1)$$

values for the concentration of water vapor in equilibrium with water (w_w) and desiccant (w_d) at the temperatures measured during the experiment are found from published data and the area of the membrane (a) determined by measurement.

TABLE 1

EFFECT OF THE SPREADING TECHNIQUES (1, 2, 3) ON THE RATE OF EVAPORATION OF WATER THROUGH STEARIC ACID MONOLAYERS

Spreading Technique →	(1)	(2)	(3)
Surface Pressure During Measurement dynes/cm		Rate of evaporation, μ g/sec	
14	433*	415*	429*
	444	426	422*
	458	435	434
	...	...	437
	427*	429*	...
	439	438	437*
		443	430*
			443
			441
Means of all readings	440	431	434
Means of best (*) values	430	422	429
22	401	402	410
	399	401	401
	403	409	401
		...	...
		390	397
		391	398
			402
Means of all readings	401	398	401

NOTE: Within each set of readings a new monolayer is indicated by three dots; readings on any one monolayer are arranged chronologically. The sequence of experiments was randomized.

(The subscripts f and w refer to measurements on a film covered surface and a clean surface respectively.)

(b) Alternatively, if the average temperatures of the water surface and of the desiccant are kept reasonably constant* equation (1) can be written:

$$r = a(w_w - w_d)(t/m_f - t/m_u) \tag{2}$$

and the value of the coefficient, $a(w_w - w_d)$, can be determined experimentally by the method suggested by Langmuir and Schaefer.[1] If the air gap (h) between the surface and the membrane is varied the following relationship holds:[2]

$$h = Da(w_u - w_d)t/m - \text{const.} \tag{3}$$

Since the diffusion coefficient (D) of water vapor in air is known (*Internat. Crit. Tables*) the coefficient $a(w_w - w_d)$ can be determined.

Previous workers[2, 3] calculated their results by method (a) whereas method (a) was used for the results reported here.

In calculating evaporation resistance it is necessary to correct the mass of water absorbed (m) for extraneous water vapor picked up from sources other than the water surface immediately beneath the membrane of the absorber box. This extraneous water vapor can come from three sources:

(a) from the air while the box is being placed in position and removed;

(b) from the air while the box is in position by diffusion through the narrow space between the surface and the supporting plate for the box;

(c) from parts of the surface not under the membrane by evaporation and diffusion to the relatively dry air under the box.

Previous workers[2, 3] with this apparatus have corrected their data for (a) and (b)

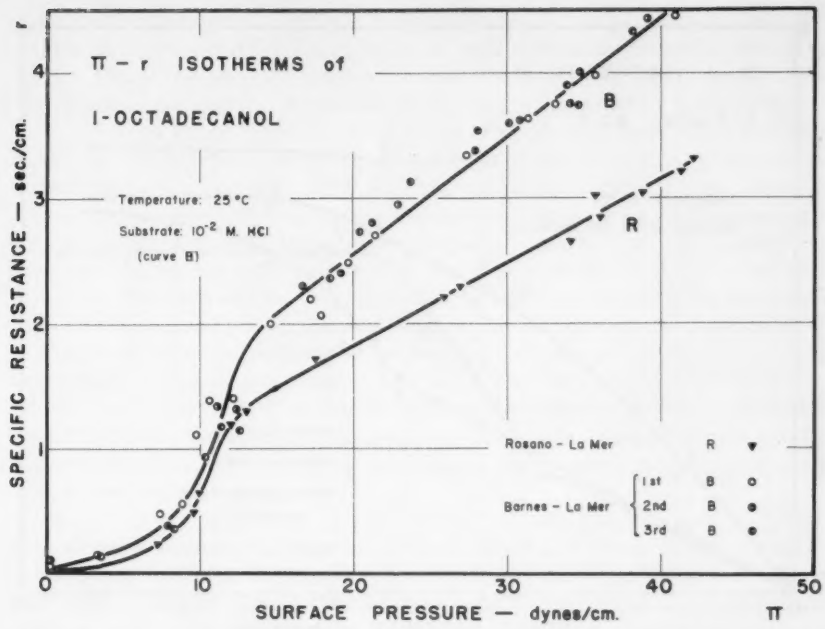

FIG. 1.—The specific evaporation resistance (r) of 1-octadecanol monolayers as a function of the surface pressure (π): determined by Rosano and La Mer[3] by technique (1) (curve R), and the present authors by technique (3) (curve B). Experimental points on curve B are differentiated to show whether they were the first, second, or third reading on a monolayer.

only, so for the purpose of comparing the present results with theirs the same type of correction has been adopted.

Discussion.—From Table 1 it is clear that we have found no significant differences between the rates of evaporation through stearic acid monolayers spread by the three different techniques. Hence, if sufficient precautions are taken, it is not necessary to use the special spreading technique of Archer and La Mer; satisfactory results can be obtained by the conventional technique (3) of spreading at zero surface pressure with a dilute solution in petroleum ether. It should be noted that monolayers spread by technique (3) are more susceptible to contamination during spreading than those spread by technique (1). It is therefore recommended that, when the conventional technique (3) is being used, a monolayer of known resistance should be employed occasionally to make sure that the system is free from contamination.

It has now been shown conclusively[2, 4, 5] that some of the spreading solvent (benzene or petroleum ether) is retained in the monolayer, the amount retained depending on the concentration of the spreading solution. We also have some evidence that the evaporation resistance can be affected by the concentration of the spreading solution at low surface pressures. Therefore, in order to minimize these solvent effects, it is desirable to use the highest concentration consistent with complete spreading (i.e., no crystals).

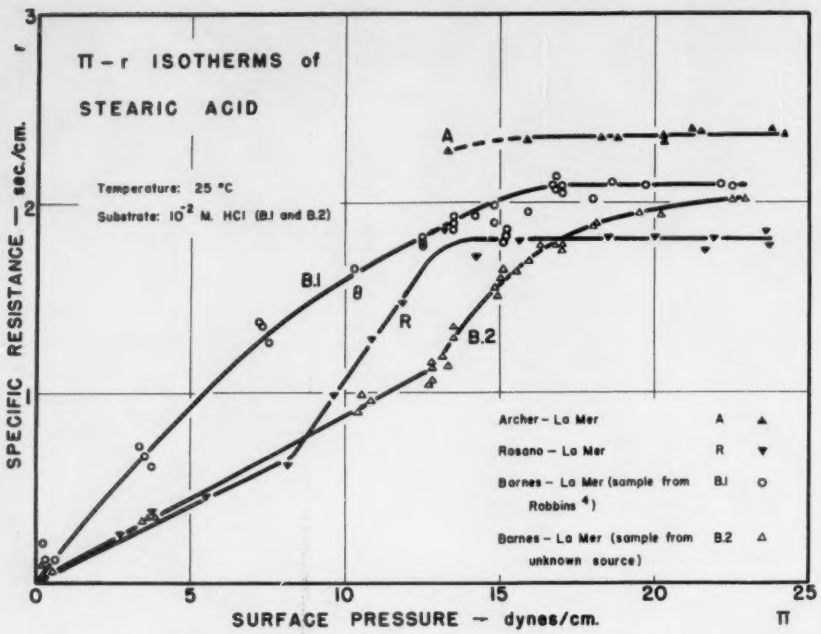

FIG. 2.—The specific evaporation resistance (r) of stearic acid monolayers as a function of the surface pressure (π): determined by Archer and La Mer[2] (curve A) and Rosano and La Mer[3] (curve R) using technique (1), and the present authors by technique (3) (curves $B.1$ and $B.2$).

The results of Table 1 also show that at the two surface pressures used, the special Archer–La Mer technique did not affect the evaporation resistance, indicating that published results using this method are reliable. This conclusion was further checked by repeating the complete π-r curves for stearic acid and octadecanol using the conventional spreading technique, and comparing the results with the published data in which technique (1) was used (Figs. 1 and 2). However, in making these comparisons it should be noted that different samples were used, whereas the results in Table 1 were all obtained with the same sample of stearic acid. In the next section we shall demonstrate the marked effect of impurities in the stearic acid samples.

With octadecanol (Fig. 1) the evaporation resistances were found to be higher than those reported by Rosano and La Mer,[3] but both π-r isotherms have the same shape. The difference probably arises from the use of different samples of 1-octadecanol, but there is also a possible source of error in the technique used by Rosano and La Mer to measure surface pressure. These authors used a Wilhelmy plate and expanded their monolayers to the desired surface pressures; a procedure which could give high readings for surface pressure due to changes in contact angle.

With stearic acid (Fig. 2) the present results (curve B.1) are intermediate between those of Archer and La Mer (curve A) and Rosano and La Mer (curve R) at surface pressures between 14 and 24 dynes/cm. Here again the differences probably arise

from the use of different samples. At lower pressures, and particularly around 10 dynes/cm the results differ considerably i.e., well outside the normal experimental error. These differences are exhibited in both the magnitude of the results and in the shapes of the curves. A possible explanation is suggested by a comparison of curves B.1 and B.2 in Figure 2. Both curves were obtained under identical conditions, the only difference being in the samples of stearic acid. Curve B.1 was obtained from a sample melting at 69.5°C, whereas curve B.2 derived from a less pure sample melting at 68.1°C. It is probable therefore that the difference between these two curves was caused by impurities in the sample used for curve B.2, and it follows that the sample of Rosano and La Mer was intermediate in purity between these two. Furthermore the shape of curves B.2 and R strongly suggests the squeezing out of an impurity from the monolayer as the surface pressure is increased.

The use of different samples and the sensitivity of the evaporation resistance to impurities in these samples makes it difficult to detect any differences in the results which could be attributed unambiguously to the different spreading techniques used. It is clear, however, that the use of the conventional spreading technique (3) has not produced any appreciable lowering in the evaporation resistances of the octadecanol and stearic acid monolayers.

The sensitivity of the evaporation resistance to contaminants poses a difficult problem in this work. The commonly used criteria of purity (such as melting point, refractive index, X-ray spacing, spectra, acid value, and iodine number) are not sensitive enough to detect amounts of impurity that could have a significant effect on the evaporation resistance. It is possible to use the evaporation resistance itself as a measure of purity, but so far there is no absolute standard by which the results can be evaluated. If we assume that the reciprocal resistances of the stearic acid (C_{18}) and a contaminant are additive[3] it can readily be shown that small amounts of palmitic acid (C_{16}) are roughly three times as effective in lowering the resistance as the same mole fraction of arachidic acid (C_{20}) is effective in raising the resistance. Considering also that the most likely contaminants in stearic acid are the lower homologues rather than the higher ones, it seems reasonable to assume that a sample of stearic acid with the highest evaporation resistance is the most pure. With this as a working hypothesis we are now studying stearic acid samples from a number of different sources, or purified by different methods, in the search for one that can be used as a reference standard.

Three conclusions can be drawn from these experiments. Firstly, if sufficient precautions are taken to prevent contamination of the monolayer, the conventional method of spreading from dilute solutions in petroleum ether at zero surface pressure is satisfactory. Secondly, the spreading method used by Archer and La Mer and by Rosano and La Mer does not vitiate their results with monolayers of one component. Thirdly, evaporation resistance is very sensitive to the presence of impurities in the sample used for the monolayer.

Our best thanks are due to Dr. R. J. Archer for valuable discussion.

This work is supported by a research grant from the National Science Foundation, to whom we are indebted. We also wish to thank the Esso Research and Engineering Company, Linden, N. J., for financial assistance in the early stages of the investigation.

* For example, variations of ±0.3° in the water temperature or ±2.5° in the desiccant temperature will each cause an error of ±2 per cent in r, or a maximum combined error of ±4 per cent.

[1] Langmuir, I., and V. J. Schaefer, *J. Franklin Inst.*, **235**, 119 (1943).
[2] Archer, R. J., and V. K. La Mer, *J. Phys. Chem.*, **59**, 200 (1955).
[3] Rosano, H. L., and V. K. La Mer, *J. Phys. Chem.*, **60**, 348 (1956).
[4] La Mer, V. K., and M. L. Robbins, *J. Phys. Chem.*, **62**, 1291 (1958).
[5] Robbins, M. L., Ph.D. dissertation, Columbia University, April, 1959.

---

*THE FORMATION OF A NEW ENZYMATICALLY ACTIVE PROTEIN AS A RESULT OF SUPPRESSION**

BY IRVING P. CRAWFORD† AND CHARLES YANOFSKY

DEPARTMENT OF BIOLOGICAL SCIENCES, STANFORD UNIVERSITY

Communicated by David M. Bonner, June 19, 1959

Suppressor mutations are changes which, though occurring at a genetic locus distinct from the site of a primary mutation, reverse the effect of the primary mutation. Depending on the nature of the primary mutation[1] there may be several means by which suppressor genes act. In certain instances the suppressor mutation may allow the synthesis of a required substance by an alternate pathway. A second possible mechanism involves relief of inhibition; if the primary mutation has resulted in the formation of an enzyme susceptible to inhibition by some normal cellular component, any mutation reducing the level of the inhibitor within the cell will, in effect, be a suppressor mutation. Examples of suppression by both the above mechanisms are known.[2–4] In addition it has been suggested that certain suppressor genes might act by permitting the formation of new, effective enzymes to function for the absent or defective ones. In *Neurospora crassa* several cases of suppression are known in mutants which form an altered protein immunologically related to the enzyme tryptophan synthetase;[5] in the presence of the suppressor gene normal enzymatic activity appears. Moreover, two instances are known in *Escherichia coli* where mutants lacking a specific enzyme or any protein immunologically related to it regain a protein with wild-type activity upon suppression.[6] In these instances, however, it has not yet been shown that a *second* protein (one the unsuppressed mutant is incapable of forming) is responsible for the activity observed, since an inhibition relief mechanism or a quantitative effect on enzyme formation might be proposed to account for the experimental findings.

In the present study examination of a suppressed mutant derived from a tryptophan-requiring strain of *E. coli* which forms a mutationally altered protein has revealed the presence of a new, enzymatically effective protein *in addition to* the altered one. This second protein is responsible for the suppressed mutant's ability to grow on unsupplemented media.

The enzyme system employed in this study, tryptophan synthetase, has been shown to catalyze the following three reactions.[7, 8]

(1) Indole + L-Serine → L-Tryptophan

(2) Indoleglycerol phosphate ⇌ Indole + Triose phosphate

(3) Indoleglycerol phosphate + L-Serine → L-Tryptophan + Triose phosphate

Of these the third is the one which appears to be responsible normally for growth in the absence of added tryptophan. The enzyme system consists of two separable protein components, termed A and B, both of which must be present for the appearance of appreciable activity in any of the three reactions.[8] The synthesis of these two proteins is controlled by separate though neighboring genes.[9] Mutationally altered forms of both components are known which, when present with the normal second component, can catalyze but one of the three reactions.[10] The mutant used in these experiments forms an altered component A which, when combined with component B, is fully active in reaction 1 but inactive in reactions 2 and 3.[10]

Since enzyme effective in permitting growth in the absence of supplement must be able to convert indoleglycerol phosphate (InGP) to tryptophan, estimates of the levels of effective and altered enzymes in an extract can be made by comparing the activities evidenced in reactions 3 and 1. Any significant deviation from a constant ratio of these two activities at different stages of purification would constitute evidence of the presence of two different enzymatically active proteins.

Materials and Methods.—The origin and certain characteristics of *E. coli* strain T–3, the mutant strain A–3 and the suppressed mutant, strain A–3 su–3, have been described previously.[10] Strain T-3 is blocked at an early step in tryptophan formation; when grown on a limiting amount of supplement it forms large amounts of the normal components A and B of tryptophan synthetase. All strains were grown on a glucose-minimal medium for extract preparation; a growth-limiting supplement of indole (2.5 μg/ml) or anthranilic acid (2 μg/ml) was added for strains A–3 and T–3 respectively. The cells were harvested at maximal growth, washed, resuspended in 0.1 *M* Tris buffer at pH 7.8, and disrupted in a 10 Kc sonic oscillator. Reversion controls were performed on aliquots of all cultures at the time of harvesting; wild-type colonies could easily be distinguished from A–3 su–3 colonies on unsupplemented minimal agar on the basis of their colony size, for after 48 hours A–3 su–3 colonies were barely visible while wild-type colonies were 2 mm in diameter or larger.

Purification and chromatography: Two procedures were used to prepare crude extracts for chromatography on DEAE cellulose. For both the first step was the precipitation of nucleic acids by the addition of a twentieth volume of 1 *M* $MnCl_2$. In method I component B was then precipitated by the slow addition of acetic acid to pH 4.0. After centrifugation the component A-containing supernatant was returned to pH 8.0 by the addition of 10 *N* NaOH, centrifuged again and brought to 60 per cent saturation with $(NH_4)_2SO_4$. The precipitate was collected by centrifugation, dissolved in a small volume of 0.01 *M* KPO_4 buffer, pH 7.2, and dialyzed against this buffer for 4–8 hours. Method 2 was used when partial purification of both component A and component B was desired. In this case the $MnCl_2$ supernatant was adjusted to pH 6.2 with acetic acid, and $(NH_4)_2SO_4$ was slowly added to 40 per cent saturation. The resulting precipitate was collected by centrifugation, dissolved in a small amount of 0.01 *M* KPO_4 buffer at pH 7.2 (containing 2 μg of pyridoxal phosphate/ml), and dialyzed 4–8 hours against this same mixture.

Following either purification method the chromatographic procedure was essentially as previously described.[8] A KCl gradient was employed for the elution of

the protein components from DEAE cellulose. When component B was present all eluting solutions contained 2 μg of pyridoxal phosphate/ml.

Assays and transduction methods: The assays for the conversion of InGP and indole to tryptophan by measuring tryptophan appearance and indole disappearance respectively have been described.[10] A unit of activity in either reaction corresponds to the conversion of 0.1 μM of reactant to product in 20 minutes at 37°C.

The transduction techniques employed, using phage P1kc, have been described elsewhere.[9]

Results.—Characterization of the suppressor mutation: Proof that strain A–3 su–3 is in fact a suppressed mutant rests on the results of transduction experiments in which strains resembling the original mutant, A–3, were recovered from it. The experiments were performed by selecting for a marker (*cys-4*) closely linked to the A–3 marker.[9] Cys$^-$ cells were infected with phage grown on strain A–3 su–3 and plated on a medium containing tryptophan but lacking cysteine. A similar experiment with phage grown on strain A–3 served as a control. The cysteine independent colonies which appeared in both transductions were picked and restreaked on test media to score the frequency of tryptophan dependence (Table 1). Approximately the same ratio of tryptophan dependent to independent colonies was

TABLE 1

TRANSDUCTION EXPERIMENTS WITH MUTANT AND SUPPRESSED MUTANT STRAINS

Donor	Recipient	Recombinants		tryp$^-$ mutants per cent
		cys$^+$tryp$^-$	cys$^+$tryp$^+$	
A–3	*cys4*$^-$	54	83	39
A–3 su–3	*cys4*$^-$	48	59	45

obtained whether strain A–3 or strain A–3 su–3 had served as the donor in this experiment. Several of the tryptophan-requiring strains recovered in this fashion from strain A–3 su–3 were tested in growth and accumulation tests and were indistinguishable from strain A–3.

Having shown that a second gene is responsible for the tryptophan independence of strain A–3 su–3, the allele specificity of this gene was examined by transduction experiments of a second type. Here strain A–3 and its allelic strains A–1 (which also forms an altered component A) and A–2 (which does not)[10] served as recipients for phage grown on strain A–3 su–3. Following adsorption, cells were plated on minimal medium. Only with A–3 cells as recipients were slow-growing, A–3 su–3-like colonies obtained in these experiments, showing that the suppressor gene under investigation is allele specific.

Comparison of normal and altered components A: In previous studies it was shown that the elution pattern of normal component A during chromatography on DEAE cellulose varies depending on the presence or absence of component B.[8] Without component B a single peak of A protein appears in the early fractions; in the presence of B, however, a biphasic distribution ensues since a portion of the A protein remains combined with B protein and is eluted in later fractions. Before examining the suppressed mutant A–3 su–3 the chromatographic behavior and properties of the abnormal component A from strain A-3 were examined. It was found that the A protein from this mutant behaves exactly like normal component A during purification. Following pH 4.0 treatment (purification method 1), and the consequent elimination of component B, this altered component is inseparable

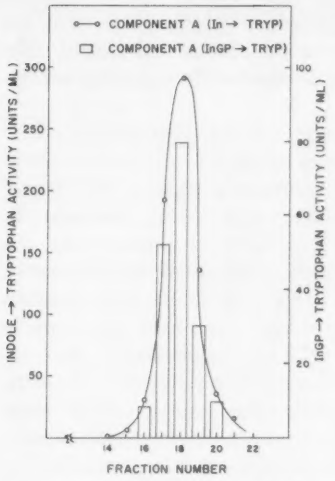
COMPONENT A (In → TRYP)
COMPONENT A (InGP → TRYP)
INDOLE → TRYPTOPHAN ACTIVITY (UNITS / ML)
InGP → TRYPTOPHAN ACTIVITY (UNITS / ML)
FRACTION NUMBER

FIG. 1.—Chromatography of a mixture of T-3 and A-3 components A. Both extracts were prepared by method 1 (see Methods). Histograms represent InGP→tryptophan activity due to normal component A. The smooth curve represents indole→tryptophan activity, the sum of the activities of normal and abnormal components A.

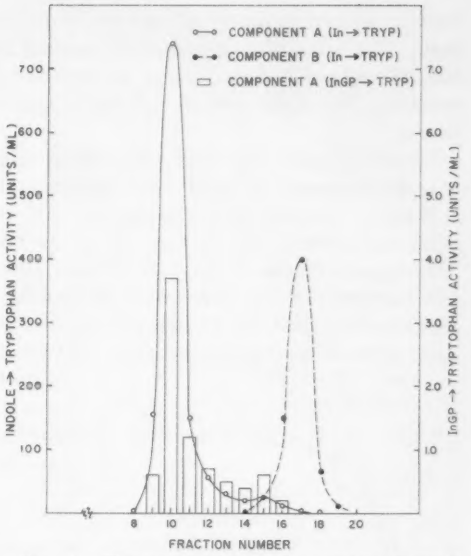
COMPONENT A (In → TRYP)
COMPONENT B (In → TRYP)
COMPONENT A (InGP → TRYP)
INDOLE → TRYPTOPHAN ACTIVITY (UNITS / ML)
InGP → TRYPTOPHAN ACTIVITY (UNITS / ML)
FRACTION NUMBER

FIG. 2.—Chromatography of an extract of strain A-3 to which a small amount of T-3 component A was added. The A-3 extract was purified by method 2 (see Methods). Histograms represent InGP→tryptophan activity due to the added normal component A. The smooth curve represents indole→tryptophan activity, the sum of the activities of normal and abnormal components A.

from normal A by chromatography on DEAE cellulose. Figure 1 shows the result of co-chromatography of the two A components purified separately and combined just prior to dialysis. Normal component A is assayed in reaction 3, while values obtained in reaction 1 represent the sum of the activities of both components. It can be seen that the normal and abnormal components behave alike under these chromatographic conditions.

Standing in contrast, however, is the behavior of altered component A when chromatographed in the *presence* of component B following purification by method 2. The biphasic distribution of A activity expected from previous experiments with normal A protein is scarcely detectable, for very little of the altered component A appears in the area of the expected "combined A–B" peak. This difference is illustrated in Figure 2, the chromatogram of an A-3 extract to which a small amount of normal component A was added just prior to chromatography. Clearly a larger proportion of the normal than the abnormal component remains associated with component B. The distribution of the normal component in this chromatogram does not seem to differ significantly from that observable in the absence of altered A protein.

On the premise that the chromatographic behavior just illustrated might result from a small difference in the affinities of normal and abnormal components A for the B protein, an investigation of the saturation of component B with each of the A proteins was undertaken. The results are shown in Figure 3; altered component A

shows a somewhat poorer affinity for component B than does its normal counterpart. The units of A required for one-half the maximal reaction rate using 2.9 and 2.8 units of B were 1.3 and 1.7 respectively for the normal and altered components. The difference, though small, was found consistently in several experiments.

Chromatography of A–3 su-3 tryptophan synthetase: As reported earlier,[10] there is a small amount of InGP → tryptophan activity present in extracts of strain A–3 su–3. As with wild-type extracts, pH 4.0 treatment inactivates the preparation, but activity can be restored by the subsequent addition of component B. This suggests that an A–like protein forms part of the effective enzyme system. An A–3 su–3 extract was purified by method 2, chromatographed in the presence of component B, and the fractions obtained assayed for both InGP → tryptophan and indole → tryptophan activity. The results of this experiment are shown in Figure 4. Though the method of representation is the same as for Figure 2, here the InGP → tryptophan activity found originated in the suppressed mutant extract itself, not in added normal component A. A definite enrichment for InGP utilizing activity in the A component of the second or "combined A–B" peak of the A–3 su-3 chromatogram is apparent. This contrasts with the results of similar experiments previously performed with *normal* tryptophan synthetase from strain T–3 where the ratio of InGP to indole utilizing activities held constant (at about 40 per cent) throughout the entire chromatogram.[8] It appears that the suppressed mutant extract contains two types of A protein. Neither in this nor in several similar experiments with A–3 su–3 extracts did more component A InGP → tryptophan activity appear in the chromatographic fractions than had been placed on the column. When assayed with normal component A, the InGP → tryptophan activity of component B in the suppressed mutant extract (not shown in Figure 4) was in constant and normal ratio to the indole → tryptophan activity throughout the B-containing area of the chromatogram, supporting the earlier finding that this component is not affected by either the original (A–3) or the suppressor (su-3) mutation.[10] The recovery of all activities in this chromatogram was in the neighborhood of 90 per cent.

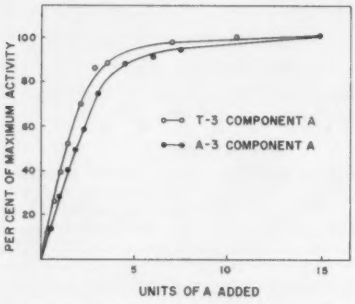

FIG. 3.—Saturation of component B with T–3 and A–3 components A. Both A components prepared by purification method 1 (see Methods). The amount of normal component B present was 2.9 units for the experiment with T–3 A and 2.8 units for the A–3 A experiment. The extent of reaction was measured in the indole→tryptophan assay.

The numerical ratio (in per cent) of the InGP → tryptophan to indole → tryptophan activity of the A–3 su–3 extract at various stages of purification is shown in Table 2, along with similar ratios from the artificial mixture experiment of Figure 2 for comparison. In each case the ratios are reasonably constant during the early stages of purification, but an enrichment of InGP utilizing activity of about 5-fold appears in the second chromatographic peak, with an inverse change in the ratios of the early fractions.

TABLE 2

RATIO OF INGP TO INDOLE UTILIZATION IN AN EXTRACT OF A-3 SU-3 AND A MIXTURE OF A-3 AND T-3 EXTRACTS AT VARIOUS PURIFICATION STAGES

Purification Stage	A-3 su-3 Extract: InGP → tryp × 100 / indole → tryp	A-3 + T-3 Extracts: InGP → tryp × 100 / indole → tryp
Crude extract	0.32	7.1
$MnCl_2$ supernatant	0.24	7.4
$(NH_4)_2SO_4$ supernatant	0.27	...
Dialysate	0.19	5.9
Chromatogram fractions		
1st peak (A alone)	0.09	3.5
2nd peak (A + B)	1.7	28

In one additional experiment the fractions in the "combined A–B" area of the chromatogram of an A-3 su-3 extract were concentrated by ammonium sulfate precipitation and rechromatographed under similar conditions. The pattern of Figure 4 reappeared, and a further enrichment of InGP → tryptophan activity (about 4-fold) was achieved in the new "combined A–B" area. The ratio of InGP to indole utilization in the best fraction still did not approach that of normal component A, however.

Examination of A-3 extracts for an inhibitor: If the A-3 mutation had resulted in the production of an A component sensitive to inhibition by some cellular constituent, the suppressor gene might act by reducing the level of the inhibitor in the cell. Under this hypothesis, the effective A component in strain A-3 su-3 would be susceptible enzyme freed of the inhibitor. Since chromatography of A-3 su-3 extracts yielded fractions with reasonable levels of InGP → tryptophan activity, it was possible to determine the susceptibility of this activity to inhibition by substances in an A-3 extract. When crude extracts of the mutant were added to suitable A-3 su-3 fractions, however, no inhibition was demonstrable (Table 3). Care was taken to maintain the usual 3 to 1 ratio of component B to component A in these assays.

TABLE 3

THE EFFECT OF CRUDE A-3 EXTRACT ON THE CONVERSION OF INGP TO TRYPTOPHAN BY A-3 SU-3 COMPONENT A

Column purified A-3 su-3 component A *ml*	Crude A-3 extract *A units*†	Partially purified component B* *units*†	InGP → tryptophan activity *units*
0.1	..	30	0.32
0.1	23	100	0.32
0.1	46	175	0.32
...	46	150	0

* This preparation of component B was free of component A.
† Determined in the indole → tryptophan reaction.

Discussion.—The results obtained in this study of suppressed mutant A-3 su-3 of *E. coli* show that in this instance suppression permits the formation of an active protein which the unsuppressed mutant cannot produce. The mechanisms of suppressor gene action established in previous studies, alternate pathway utilization and relief of inhibition, appear unlikely in this case. The suppressed strain A-3 su-3 can form tryptophan in the usual way, by the conversion of InGP to tryptophan. The inhibition release explanation also seems ruled out not only by the fact that crude mutant extract does not impair the InGP → tryptophan activity of

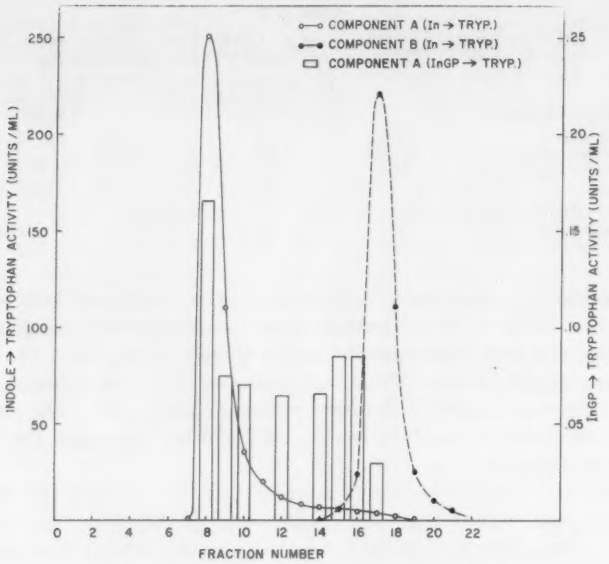

FIG. 4.—Chromatography of an extract of strain A–3 su–3 following purification by method 2 (see Methods). The method of representation is the same as in Figure 2. Fractions 7–17 were concentrated 20-fold by $(NH_4)_2SO_4$ precipitation prior to assay.

enzyme from the suppressed mutant, but by the fact that a high degree of purification, such as is attained by the procedures used,[8] does not result in either the appearance of InGP utilization in the mutant extract or the liberation of *additional* InGP utilizing ability in the suppressed mutant extract. On other grounds the likelihood of an inhibitor being closely bound to the "altered" A protein seems remote since this protein (in the presence of component B) has the same activity to antigen ratio in the indole → tryptophan reaction as does normal component A.[10] Thus the hypothetical inhibitor would have to eliminate completely one reaction catalyzed by the protein without having any effect on its participation in a second reaction.

The possibility that the gene responsible for suppression in strain A–3 su–3 can cause the *de novo* synthesis of a protein which can serve as an effective component A was ruled out in this case by the finding that the suppressor gene in question is specific for the A–3 allele and does not alleviate the tryptophan requirement of other A alleles.

It remains possible that the original mutant, A–3, is potentially capable of producing effective component A in addition to the altered protein, and that the effect of the suppressor gene is to relieve an inhibition of this synthetic ability. At the present time, however, it seems more likely that the effect of the suppressor gene is to modify somehow a small portion of the product of the mutant gene. Such an interpretation is not only in accord with the experimental data but forms a close analogy with results recently obtained in allele heterocaryon complementation studies in *Neurospora crassa*.[11] Assuming, as seems likely from their similar proper-

ties, that the normal and mutationally altered component A proteins differ but slightly, it is possible that in the presence of the mutant form of the suppressor gene a short amino acid sequence similar to that of the affected area of component A might be exchanged with the altered region. Any mechanism (analogous to transpeptidation if operative at the protein level but possibly occurring at an earlier stage of protein synthesis) permitting exchange in this manner might allow the elaboration of a small amount of effective enzyme. A somewhat similar mechanism has been postulated to explain the ability of certain pairs of alleles in *N. crassa* to form "normal" enzyme in a heterocaryon.[11]

Whatever the mechanism by which the effective component A in strain A–3 su–3 is formed, it is clear that the distribution of this protein in chromatograms where component B is present is very similar to that of normal, mutationally unaltered component A under similar conditions. The present experiments do *not*, however, permit the conclusion that the effective protein *is* normal component A, in the sense of being identical with the protein present in wild type cells. Analogous phenomena may account for several unexplained instances of suppression at other loci and in other organisms.[12–15] A determination of the prevalence of the mechanism of suppression uncovered in these experiments with *E. coli* strain A–3 su–3 must await the successful application to other systems of more refined protein separation and identification techniques, however.

Summary.—Certain tryptophan auxotrophs of *E. coli* form a defective A protein, one of the components of the tryptophan synthetase enzyme system. Examination of a suppressed mutant of one such strain has revealed that, in addition to the altered protein characteristic of the unsuppressed mutant, an A protein with normal catalytic activity is present. Hypotheses for the formation of the second A protein are discussed.

The authors are greatly indebted to Mrs. Virginia Horn and Mrs. Mary Hurd for technical assistance.

* This investigation was supported by grants from the National Science Foundation and the U. S. Public Health Service (RG-5652).

† Postdoctoral fellow of the National Foundation.

[1] In this paper we confine discussion to primary mutations which exert a direct effect on an enzymatically active protein. If instead, for instance, the primary mutation were to cause the synthesis of an inhibitory substance, other forms of suppression would be possible.

[2] Suskind, S. R., and L. I. Kurek, these PROCEEDINGS, **45**, 193 (1959).

[3] Lein, J., and P. S. Lein, these PROCEEDINGS, **38**, 44 (1952).

[4] Strauss, B. S., and S. Pierog, *J. Gen. Microbiol.*, **10**, 221 (1954).

[5] Yanofsky, C., in *Enzymes, Units of Biological Structure and Function*, ed. O. H. Gaebler (New York: Academic Press, 1956), p. 147.

[6] Yanofsky, C., *Science*, **128**, 843 (1958).

[7] Yanofsky, C., *Biochem. Biophys. Acta*, **31**, 408 (1959).

[8] Crawford, I. P., and C. Yanofsky, these PROCEEDINGS, **44**, 1161 (1958).

[9] Yanofsky, C., and E. Lennox, *Virology* (in press).

[10] Yanofsky, C., and I. P. Crawford, these PROCEEDINGS, **45** (in press).

[11] Woodward, D. O., C. W. H. Partridge, and N. H. Giles, these PROCEEDINGS, **44**, 1237 (1958).

[12] Yura, T., *Genetic Studies with Bacteria*, Carnegie Inst. of Wash. Publ. **612**, 77 (1956).

[13] Starlinger, V. P., and F. Kaudewitz, 3. *Naturforschg.*, **116**, 317 (1958).

[14] Yanofsky, C., and D. M. Bonner, *Genetics*, **40**, 761 (1955).

[15] Giles, N. H., and C. W. H. Partridge, these PROCEEDINGS, **39**, 479 (1953).

A MATHEMATICAL THEORY OF ADAPTIVE CONTROL PROCESSES

By Richard Bellman and Robert Kalaba

THE RAND CORPORATION, SANTA MONICA, CALIFORNIA

Communicated by Philip M. Morse, June 26, 1959

Introduction.—In the last few years, the mathematical theory of control processes has attracted a great deal of attention.[1–7] Problems of stability and optimization have been posed and resolved for deterministic and stochastic processes. Along with this intensive analytic effort has come a more searching examination of the relevance of existing mathematical models of the underlying physical processes.[8, 9] It is now recognized that questions of engineering feedback and industrial process control, design of automata, control of man-machine combinations, data-processing on a large scale, and many other activities as well, contain elements of uncertainty of various unconventional types which escape classical formulation and classical treatment. The processes that occur are "learning"[10–12] or "adaptive" processes in which it is required to act and learn simultaneously. Although some preliminary work has been done from the statistical side[13–15] the foundations for a general theory have not as yet been laid.

The purpose of this note is to indicate how a general mathematical framework can be constructed using the techniques of the theory of dynamic programming.[16] More detailed accounts of particular applications will appear in subsequent publications;[17, 18] see also Bellman and Kalaba,[19] and Bellman.[20]

Formulation.—In a deterministic process, the state of the system is specified by a point in phase space, p, and the transformations, $T(p, q)$, resulting from decisions q, yield points of the same nature. In stochastic processes, the state of the system is specified by coordinates of this type, and in addition, by probability distributions, "points" in a "phase space" of more general type. At a sufficient level of abstruction, these are projections of a general multi-stage decision process.

Introducing a further level of abstraction, we shall show how the same conceptual and analytic techniques, and computational methods as well, can be applied to the study of adaptive control processes.

Let the state of the system S be specified, as usual, by a point p in phase space, and, what is new, by an *information pattern*, P. This information pattern represents the information about the process that we wish to retain in order to guide our further actions. As a consequence of a decision, q, p is transformed into a new point p_1, given *a priori* by the transformation $p_1 = T_1(p, P; q, r)$, and P is transformed into a new information pattern, given *a priori* by the transformation $P_1 = T_2(p, P; q, r)$. Here r is a random vector variable, specified by an *a priori* probability distribution $dG(p, P; q, r)$, itself a part of the information pattern P.

After the decision has been made, p_1 may or may not be completely determined. Let us assume here, for the sake of simplicity, that these are known after the decision has been made. We shall suppose also that the transformations T_1 and T_2 are known, although in many adaptive processes, the determination of these functions, and even of the state and information patterns themselves, is an essential part of the problem.

With this format, let it be required to determine a sequence of decisions, $[q_1, q_2,$

. . . , q_N], which will minimize the expected value of a preassigned function of the terminal state, (p_N, P_N). The expected value is taken with respect to the set of *a priori* probability distributions.

Functional Equations.—Following the usual approach of dynamic programming, we introduce the sequence of functions

$$f_N(p, P) = \text{Min Exp } \phi(p_N, P_N) \tag{1}$$

Then the principle of optimality yields the functional equation

$$f_N(p, P) = \min_{q_1} [\int f_{N-1}(T_1(p, P; q, r), T_2(p, P; q, r))dG(p, P; q, r)] \tag{2}$$

for $N = 2, 3, \ldots$, with

$$f_1(p, P) = \min_{q} [\int \phi(T_1(p, P; q, r), T_2(p, P; q, r))dG(p, P; q, r) \tag{3}$$

These relations can be used to establish the existence of optimal policies, and to obtain various structural characteristics, as in Bellman.[16]

Sufficient Statistics.—For analytic and computational purposes, it is desirable to replace the sequence $f_N(p, P)$, which may be a sequence of functionals due to the dependence upon P, by a sequence of functions. A number of devices, familiar in modern mathematical statistics, and collected under the heading of "sufficient statistics," permit this very important reduction. See[3, 21, 22] for applications of this method, and for further discussion of particular problems.

Discussion.—In view of the very uncertainty that exists as to how to treat uncertainty, there can be no pretense of ever erecting a definitive theory of adaptive processes. The erection of such a theory is itself a sequential learning process. What we wish to construct is a theory which is plausible, feasible and flexible, and which can be shown, under certain favorable circumstances, to converge over time, with probability one in any particular process, to the theory which we possess for treating deterministic and stochastic control processes. Matters of this nature will be discussed in detail in subsequent publications.

[1] Truxal, J., *Automatic Feedback Control System Synthesis* (McGraw-Hill, 1955).

[2] Laning, J., and R. Battin, *Random Processes in Automatic Control* (McGraw-Hill, 1956).

[3] Wiener, N., *Nonlinear Problems in Random Theory* (John Wiley and Sons, 1958).

[4] Bellman, R., "Dynamic Programming and the Computational Solution of Feedback Design Control Processes," AIEE Special Publication T-101, *Proc. Computers in Control Systems Conference*, 1957.

[5] Bellman, R., "Dynamic Programming and Stochastic Control Processes," *Information and Control*, **1**, 228–239 (1958).

[6] Bellman, R., and R. Kalaba, "On adaptive control processes," 1959 IRE National Convention Record (to appear).

[7] LaSalle, J. P., "Time Optimal Control Systems," these PROCEEDINGS, **45**, 573–577 (1959).

[8] Zadeh, L., "What Is Optimal?", *IRE Trans. on Information Theory*, **IT–4**, 3 (1958).

[9] McMillan, B., "Where Do We Stand?", *IRE Trans. on Information Theory*, **IT–3**, 173–4 (1957).

[10] Bush, R. R., and F. Mosteller, *Stochastic Models for Learning* (John Wiley and Sons, 1955).

[11] Bellman, R., T. E. Harris, and H. N. Shapiro, *Studies in Functional Equations Occurring in Decision Processes*, RM-878 (Rand, 1952).

[12] Flood, M. M., *On Game–Learning Theory and Some Decision–Making Experiments, Decision Processes* (John Wiley and Sons, 1954).

[13] Wald, H., *Sequential Analysis*, (John Wiley and Sons, 1947).

[14] Robbins, H., "Some Aspects of the Sequential Design of Experiments," *Bull. Amer. Math. Soc.*, **58**, 527–535 (1952).

[15] Thompson, W. R., "On the Theory of Apportionment," *Amer. Jour. Math.*, 450–457 (1935).

[16] Bellman, R., *Dynamic Programming* (Princeton University Press, 1957).

[17] Bellman, R., and R. Kalaba, "On Adaptive Control Processes," 1959 IRE National Convention Record, (to appear).

[18] Bellman, R., and R. Kalaba, *Dynamic Programming and Adaptive Processes I: Mathematical Foundation*, P-1416, (Rand, 1959).

[19] Bellman, R., and R. Kalaba, "On Communication Processes Involving Learning and Random Duration," 1958 IRE National Convention Record, Part 4, 1958.

[20] Bellman, R., "A Problem in the Sequential Design of Experiments," *Sankhya*, **16**, 221–229 (1956).

[21] Aoki, M., Ph.D. thesis, UCLA (to appear).

[22] Freimer, M., "A Dynamic Programming Approach to Adaptive Control Processes," Lincoln Laboratory (1959).

PARTITIONS IN HOMOGENEOUS, FINITE ABELIAN GROUPS

BY ECKFORD COHEN

UNIVERSITY OF TENNESSEE

Communicated by Jesse Douglas, June 15, 1959

Let r represent a positive integer. A (complex-valued) arithmetical function $f_r(n)$ is called *even* (mod r) if $f_r(n) = f_r((n, r))$ for all integers n, where (n, r) denotes the greatest common divisor of n and r. Since the integers (mod r) form an additive cyclic group, one may rephrase this definition as follows. Let C_r be a cyclic group of order r with elements α. A function $f_r(\alpha)$ defined in C_r will be termed even if $f_r(\alpha)$ is invariant under all automorphisms of C_r.

The notion of even function (mod r) has been extended to functions of several variables in a paper as yet unpublished. In that paper a function $f_r(n_1, \ldots, n_k)$ of k integral variables is defined to be (relatively) *even* (mod r) if $f_r(n_1, \ldots, n_k) = f_r((n_1, r), \ldots, (n_k, r))$ for all n_i; if, moreover, there exists an even function $F_r(n)$ such that $f_r(n_1, \ldots, n_k) = F_r((n_1, \ldots, n_k))$ for all n_i, then $f_r(n_1, \ldots, n_k)$ is defined to be *totally even* (mod r).

In the present note we restate in group-theoretical terminology some of the concepts and results of the above cited paper. In particular, one may reformulate the preceding definitions as follows. Let $G_r^{(k)}$ denote an (additive) abelian group, decomposable into a direct sum of k cyclic groups, $C_r^{(i)}$, each of order r,

$$G_r^{(k)} = C_r^{(1)} \oplus \ldots \oplus C_r^{(k)}. \tag{1}$$

Let $f_r(\alpha)$ be a complex-valued function defined in $G_r^{(k)}$. We shall say that $f_r(\alpha)$ is *relatively even*, or more precisely, even relative to the decomposition (1), if $f_r(\alpha)$ is invariant under all automorphisms of $G_r^{(k)}$ which induce automorphisms in $C_r^{(i)}$, $i = 1, \ldots, k$. The function $f_r(\alpha)$ will be termed *totally even* if it is invariant under the totality of automorphisms of $G_r^{(k)}$. The equivalence of the latter definition with the definition of totally even function (mod r) stated above is a consequence of the fact that, in the homogeneous group $G_r^{(k)}$, each cyclic subgroup of order r is a direct summand.[1]

Define[2] $J_k(r)$ to be the number of elements of order r in $G_r^{(k)}$ and place $\phi(r) = J_1(r)$, $J(r) = J_2(r)$. We next state some results on binary compositions in abelian groups which are reformulations of results proved in the paper mentioned above. Let us specialize to the case $k = 2$ in (1), placing $G_r = G_r^{(2)} = C_r^{(1)} \oplus C_r^{(2)}$. The indices of the components of α in $C_r^{(1)}$ and $C_r^{(2)}$ will be denoted by m and n, respectively.

We define functions $\theta_r(\alpha)$, $\omega_r(\alpha)$, and $K_r(\alpha)$ to be the number of ordered pairs ξ, η satisfying the equation

$$\alpha = \xi + \eta \qquad (\xi, \eta \epsilon G_r), \tag{2}$$

under the following sets of conditions[3]; (*i*) $\theta_r(\alpha)$ represents the number of solutions of (2) in elements ξ, η of order r; (*ii*) $\omega_r(\alpha)$ denotes the number of solutions of (2) in elements ξ, η of order r such that the components of ξ are generators of $C_r^{(1)}$ and $C_r^{(2)}$; (*iii*) $K_r(\alpha)$ represents the number of solutions ξ, η of (2) such that the subgroup generated by each pair ξ, η has exponent r.

In particular we have the following results: *The function $\omega_r(\alpha)$ is relatively even and has the explicit evaluation,*

$$\omega_r(\alpha) = \phi^2(r) \sum_{\substack{d|r \\ (d,\, mn) = 1}} \frac{\mu(d)}{\phi^2(d)}, \tag{3}$$

where $\mu(r)$ denotes the Möbius function; the functions $\theta_r(\alpha)$ and $K_r(\alpha)$ are totally even and have the evaluations,

$$\theta_r(\alpha) = J(r) \sum_{\substack{d|r \\ (d,\, m,\, n) = 1}} \frac{\mu(d)}{\phi^2(d)} \tag{4}$$

$$K_r(\alpha) = r^2 \sum_{d|(m,\, n,\, r)} \frac{\mu(d)}{d^2} = \left(\frac{r}{(m,\, n,\, r)}\right)^2 J((m,\, n,\, r)). \tag{5}$$

On the basis of these formulas it is not difficult to deduce the following solvability criterion:

For all r and all elements α of G_r, $\theta_r(\alpha) > 0$, $K_r(\alpha) > 0$; moreover, $\omega_r(\alpha) = 0$ if and only if r is even and m and n are both odd.

[1] Szele, T., *Commentarii Mathematici Helvetici*, **22**, 117–124 (1949), in particular, Satz 1.

[2] This group-theoretical definition of the Jordan totient is due to Miller, G. A., *Amer. Math. Monthly*, **11**, 129–130 (1904).

[3] In arithmetical form, (2) becomes the pair of congruences, $a \equiv x_1 + y_1$, $b \equiv x_2 + y_2 \pmod r$, $(a, r) = m$, $(b, r) = n$, and the three sets of conditions become respectively, (i) $(x_1, x_2, r) = (y_1, y_2, r) = 1$, (ii) $(x_1, r) = (x_2, r) = (y_1, y_2, r) = 1$, (iii) $(x_1, x_2, y_2, y_2, r) = 1$.

ON ENTIRE SOLUTIONS OF NONLINEAR DIFFERENTIAL EQUATIONS*

By Raymond Redheffer

UNIVERSITY OF CALIFORNIA AT LOS ANGELES

Communicated by Einar Hille, June 11, 1959

Some time ago[9] the author gave a systematic method of proving Liouville theorems and, thereby, theorems about entire solutions. This method has been applied to the equation

$$\Delta u \geq f(u, u^*), \qquad u^* = |\operatorname{grad} u| \tag{1}$$

with results which are thought to be worth summarizing here.

If $r = (\Sigma x_i^2)^{1/2}$ we set

$$v_c = c \log\log r \qquad \text{or} \qquad v_c = \int_b^r [r \log (c/r)]^{1-n} dr$$

for $n = 2$ or $n > 2$ respectively. The maximum principle applied to $u - v_c$ now yields:

THEOREM 1. *Let u be a differentiable function in the whole space, with* $\sup u = m$. *Suppose there is a positive ϵ such that*

$$\Delta u \geq -(1/\epsilon)(u^*)^2 \qquad (n = 2)$$

$$\Delta u \geq \epsilon (u^*)^{n/(n-1)} \qquad (n \geq 3)$$

in the x set where simultaneously

$$m - \epsilon < u < m, \qquad 0 < u^* < \epsilon.$$

Then $u \equiv m$.

Since (1) goes into an equation of the same form for v, if $u = y(v)$, a Liouville theorem for v can give rise to a theorem about entire solutions, u. We give the result for arbitrary n; the analog for $n = 2$ does not require $y' \leq A f(y, y')$:

THEOREM 2. *Let $f(u, p)/p^{n/(n-1)} \equiv g(u, p)$ be a nonincreasing funtion of $p > 0$ for all sufficiently large u, but positive for all u. Suppose further that $g(u, p)$ is bounded away from 0 when u ranges over compact sets and p is small. If there is a constant A such that the problem*

$$y'' \leq A f(y, y'), \qquad y' \leq A f(y, y') \tag{2}$$

has a solution with $y' > 0$, $y \to \infty$ as $x \to 1-$, then every entire solution of

$$\Delta u \geq f(u, u^*), \qquad n \geq 2$$

is constant.

To illustrate the use of this result we say that $f(u)$ has *Property* P_1 if there is a constant $\alpha > 0$ such that

$$f(u) \geq \frac{\alpha}{u} \int_0^u f(t)dt \qquad \text{when } u \gg 1.$$

The function has *Property* P_2 if the same is true, with the inequality reversed. We define

$$F(y) = \int_0^y f(t)dt, \qquad G(p) = \int_1^p \frac{p}{g(p)}\,dp$$

and state:

THEOREM 3. *Let $f(u)$ be continuous and positive for all u, and have Property P_1. Suppose $g(p)\ p^{n/(1-n)}$ is continuous, positive, and nonincreasing for $p > 0$, and is bounded away from 0 for p small. Let $p/g(p)$ have Property P_2. If*

$$\int^{\infty} \frac{dy}{G^{-1}[F(y)]} < \infty \tag{3}$$

then every entire solution of $\Delta u \geq f(u)\ g(u^*)$ is constant.

Since increasing functions have Property P_1 this contains the main results of the papers in references [1–7] and [10]; we just take $g \equiv 1$.

If

$$\int_{-\infty}^{y} f(t)dt < \infty$$

it is clear that the divergence of the integral (1) is both necessary and sufficient for a nonconstant entire solution.

When $n = 2$, Properties P_1 and P_2 are not necessary; we can take $f(y) \geq 0$ rather than > 0; and instead of continuity, we need f and g to be the derivatives of their integrals.

Although the exponents 2 and $n/(n-1)$ in THEOREM 1 cannot be improved, the statement itself can. Let

$$\Delta u \geq -(u^*)^2 h(u^*)$$

in the set $0 < u^* < \epsilon$, where $h(p)\ \epsilon C^{(1)}$, $h(p) > 0$, $h(p)(-\log p)^a$ decreases for some constant $a < 1$, and $h(p)(-\log p)^{-b}$ increases for some constant b. (These hypotheses are required when $p > 0$.) If $n = 2$ and

$$\int_0 \frac{dp}{h(p)\ p \log(1/p)} = \infty$$

then u satisfies the Liouville theorem, whereas if the integral converges there is a nonconstant entire solution u satisfying $|u| < \epsilon$, $u^* < \epsilon$. Similarly, let

$$\Delta u \geq (u^*)^{n/(n-1)}\ h(u^*) \tag{4}$$

in the set $0 < u^* < \epsilon$, where $h(p)\epsilon C^{(1)}$, $h(p) > 0$, $h(p)$ is increasing, and $p^{-a}h(p)$ is decreasing for some constant a. (These conditions are needed for $p > 0$.) If $n \geq 3$ the same conclusion holds as for $n = 2$ above, the criterion being the divergence or convergence of

$$\int_0 \frac{h(p)}{p}\,dp\ . \tag{5}$$

If (4) is required for all u^*, not only for $u^* < \epsilon$, and if the integral (5) diverges, then every entire solution u is bounded, hence constant.

* The preparation of this paper was sponsored by the Office of Naval Research. Reproduction in whole or in part is permitted for any purpose of the United States Government.

[1] Haviland, E. K., "A note on unrestricted solutions of the differential equation $\Delta u = f(u)$," *J. London Math. Soc.*, **26**, 210–214 (1951).

[2] Keller, J. B., "Electrohydrodynamics I. The equilibrium of a charged gas in a container," *J. Rational Mechanics and Analysis*, **5**, 715–724 (1956).

[3] Redheffer, R., "On entire solutions of nonlinear equations" (Abstract), *Bull. Amer. Math. Soc.*, **62**, 408 (1956).

[4] Walter, W., "Über ganze Lösungen der Differentialgleichung $\Delta u = f(u)$", *Jber. Deutsch. Math. Verein*, **57**, 94–102 (1955).

[5] Wittich, H., "Ganze Lösungen der Differentialgleichung $\Delta u = e^u$," *Math. Z.*, **49**, 579–582 (1944).

[6] Keller, J. B., "On solutions of $\Delta u = f(u)$," *Comm. Pure Appl. Math.*, **10**, 503–510 (1957).

[7] Calabi, E., "An extension of E. Hopf's maximum principle," *Duke Math. J.*, **25**, 45–56 (1957), (Thm. 4).

[8] Redheffer, R., "Maximum principles and duality," *Monatshefte f. Math.* **62**, 56–75 (1958).

[9] *Ibid.*, "A general maximum modulus theorem," NAR Report, Univ. of Calif., Los Angeles (January 1956).

[10] Osserman, R., "On the inequality $\Delta u \geq f(u)$," *Pac. J. Math.*, **7**, 1641–1647 (1957).

---

*CRITICAL SIGNIFICANCE OF BREAST STRUCTURE IN THE INDUCTION OF MAMMARY CANCER IN THE RAT**

BY CHARLES HUGGINS, LORRAINE C. GRAND, AND FILOMENA P. BRILLANTES

THE BEN MAY LABORATORY FOR CANCER RESEARCH, THE UNIVERSITY OF CHICAGO

Communicated June 29, 1959

When maximal amounts of 3-methylcholanthrene[1] compatible with survival are fed in single or multiple doses to young adult female rats of a special strain, mammary cancer develops selectively, often exclusively, with great frequency and in a short time.[2] In this paper, it will be demonstrated that the functional state of the mammary substratum exposed to a carcinogen is of critical significance in determining whether or not mammary cancer will develop. In the following experiments in which 3-MC was fed to rats it was found that mammary cancer did not develop in the atrophic mammary tubules of hypophysectomized animals and that the incidence of cancer of the breast was reduced, as well, in very hyperplastic mammary glands, so induced by the administration of hormones: the controls invariably developed mammary cancer.

The induction of mammary cancer following administration of 3-MC at some distance from the mammary gland was discovered by Maisin and Coolen[3] who found that painting the skin of mice with this cancer-producing substance often induced breast cancer in addition to cancer of the skin. Subsequent investigations revealed that this phenomenon is evoked in mice and rats by other aromatic carcinogens[4, 5] which can be successfully administered by various routes. The Maisin effect is dependent on the hormonal status of the rodent treated with the mammary carcinogen. In carcinogen-fed rodents, mammary cancer occurs in males uncommonly[6] and the incidence is low in female rats treated with testosterone,[7] other androstane derivatives[2] and estradiol-17β;[7] the incidence of mammary tumors is enhanced by treatment with progesterone.[2, 8]

Many mammary carcinomas of rats evoked by 3-MC feeding are hormone-de-

pendent since they shrink profoundly[2] after appropriate modifications of the endocrine status of the host of the established neoplasm; such hormonal dependence is a unique property among experimental breast cancers available at the present time.

Experimental.—Albino rats of the Sprague-Dawley strain were obtained from the dealer at age 42 days and kept thereafter in a controlled climate. Infectious disease was not detected in the animals during the experiments. All operations were carried out under ether anesthesia; ovariectomy or adrenalectomy was performed at age 42 days and hypophysectomy at 49 days. Hypophysectomized rats were maintained on a synthetic, high protein (18 per cent) diet consumed *as libitum;* other rats were fed a commercial ration.[9] Adrenalectomized rats were maintained by furnishing saline as drinking fluid and steroids were not administered to them. Control rats were always intact young females from the same batch and of the same age as the experimental groups.

Throughout this paper, dosage refers to the amount of compound administered each day. 3-MC, 1 gm. was dissolved in sesame oil, 100 ml, by heating briefly at 100°C. This compound, 10 mg, was administered always by gastric instillation. This dosage of 3-MC was given to hypophysectomized or adrenalectomized rats only once each week because of their decreased tolerance of the compound; their controls were treated similarly. All other rats received 3-MC, 10 mg, 6 times each week. Experiments on hypophysectomized rats started at age 63 days; other experiments began at age 49 days.

Steroids were dissolved in ethyl alcohol which was diluted with sesame oil to make the final alcoholic concentration 10 per cent; the solution (0.2 ml) was injected 6 days each week in a hind-leg. Equine gonadotrophin[10] derived from the serum of pregnant mares was dissolved in saline (0.5 ml) and injected subcutaneously in the scruff of the neck; dosage of equine gonadotrophin is expressed in the units of Cartland and Nelson.[11] Rats to be subjected to intense estrogenic stimulation (injected with estradiol-17β, 20μg or more, or with gonadotrophin) underwent hysterectomy at age 42 days.

The rats were observed until mammary cancer had developed or until 300 days had elapsed. The time of appearance of mammary cancer is dated from the first dose of 3-MC. The animals were examined for tumors by palpation daily. Histologic sections of many of the tumors were studied.

Results.—It was found that an experienced examiner could detect by palpation mammary cancer as small as 1 mm in diameter in the rat; subsequent growth of the tumor facilitated its recognition but histologic verification was essential for diagnosis.

There were 314 animals in 30 consecutive series of intact adult females fed 3-MC, 10 mg, 6 days each week, used for controls. Every one of these rats developed mammary carcinoma, and all within 92 days (Table 1). Among these 314 rats, there were 14 animals (4.46%) which developed mammary cancer after 10 weeks. All of the mammary tumors were classified as carcinoma. The next commonest tumor in these animals was parotid carcinoma observed in 17 animals in a total of 1186 rats. Parotid cancer arose in 209 $\pm$ 70 days, considerably later than mammary cancer.

Effects of Ovariectomy, Adrenalectomy, and Hypophysectomy.—Ovariectomy lowered the incidence of mammary cancer in rats fed 3-MC to 68.5 per cent (Table

TABLE 1

INFLUENCE OF EXTIRPATION OF ENDOCRINE GLANDS ON INCIDENCE AND APPEARANCE OF MAMMARY CANCER

All rats received 3-methylcholanthrene, 10 mg, by intragastric instillation according to the given schedule. The control rats were unoperated females of the same age as the experimental animals. The rates were observed for 10 months.

Endocrine Status	3-MC: Doses per Week	No. Rats	Rats with Cancer	Time of Apperarance of Mammary Cancer, Days: Range	Median	Mean
Intact controls	6	314	314	29–92	53	53.1 ± 9.4
Ovariectomy	6	108	74	45–271	81	93.4 ± 45.1
Intact controls	1	9	9	35–83	62	60.4 ± 15.9
Adrenalectomy	1	5	5	49–81	69	64.8 ± 12.8
Intact controls	1	9	9	35–83	62	60.4 ± 15.9
Hypophysectomy	1	13	0	No mammary cancer		

±, Standard deviation.

1). But it is noteworthy that in 45 ovariectomized rats (41.7%) in the series of 108 animals (Table 1) mammary cancer was detected before 90 days. Bilateral adrenalectomy, with maintenance of life on saline, failed to retard (Table 1) the induction of mammary cancer; regeneration of adrenal glands was not observed in these rats.

Hypophysectomized rats live for prolonged periods when fed a high protein diet.[12] Thirteen hypophysectomized rats survived more than 10 months after treatment with 3-MC, 10 mg, once each week for 7 weeks (total dose, 80 mg); no neoplasms of any sort were observed in any rat while all of the controls (Table 1) developed mammary cancer.

Effect of Equine Gonadotrophin.—Each member of a series of 15 rats injected with equine gonadotrophin, 0.1 unit, and fed 3-MC, 10 mg, developed mammary cancer (Table 2); the time of appearance of the neoplasm was slightly delayed as compared with the control series. Other tumors observed in this series were 1 fibroadenoma of the breast and 1 sarcoma at the injection site.

Subcutaneous injection of equine gonadotrophin, 0.4 unit, greatly delayed or prevented the appearance of mammary cancer in rats fed 3-MC (Table 2). Five series comprising 53 rats received this dosage for 50, 104, or 150 days. In these rats mammary carcinoma was observed in 13 animals (24.5%); the time of appearance was 79 to 273 days; median 150 days, mean 152.5 ± 51 days. Other mammary tumors observed were fibroadenoma in 2 rats, at 249 and 259 days, respectively. All rats which received equine gonadotrophin, 0.4 unit, for 150 days, developed a sarcoma in the subcutaneous tissue at the injection site (Table 2). The sarcomas were excised under ether anesthesia when they were detected.

TABLE 2

EFFECT OF GONADOTROPHIN ON INDUCTION OF MAMMARY CARCINOMA AND SUBCUTANEOUS SARCOMA

Gonadotrophin was injected subcutaneously, 6 times weekly, in hysterectomized rats; the control rats were intact females of the same age. Both groups received 3-methylcholanthrene, 10 mg-6 times a week for 50 days by gastric instillation. The rats were observed for 300 days.

Dose of Gonadotrophin	No. Rats	No. with Mammary Cancer	Occurrence of Mammary Cancer, Days	No.	Sarcoma at Injection Site, Time of Appearance, Days
None; controls	10	10	55.1 ± 9.4	0	..
Gonadotrophin, 0.4 units	15	3	109, 209, 273	15	217 ± 26.9
Gonadotrophin 0.1 units	15	15	72.2 ± 20.8	1	168 days

±, Standard deviation.

TABLE 3

ENDOCRINE EFFECTS OF HORMONES ADMINISTERED WITH 3-METHYLCHOLANTHRENE

Compounds were injected daily into intact female rats from age 49 to 99 days; 3-methylcholanthrene, 10 mg, was fed 6 times a week to every rat for the same period. There were 6 rats in each group.

Compound Injected	Final Body Weight, gm	Δ, gm	Pituitary, mg	Adrenals, mg	Ovaries, mg	Uterus, mg
None; uninjected	225	73	11.9 ± 0.6	73 (62–88)	78 (65–93)	520 (455–610)
Progesterone, 4 mg	271	104	12.1 ± 1.8	68 (59–75)	43 (37–52)	252 (180–310)
BKP, 0.25 mg	256	111	11.0 ± 1.9	60 (53–71)	29 (25–38)	214 (183–255)
BKP, 0.5 mg plus estradiol-17β, 2 μg	231	86	14.2 ± 1.7	64 (51–74)	29 (26–37)	347 (282–408)
Equine gonadotrophin 0.4 units	220	67	13.8 ± 1.0	58 (53.8–62.4)	264 (146–398)	...

±, Standard deviation.
Δ, Gain in weight during the experiment.

When administered concomitantly with 3-MC feeding, equine gonadotrophin, 0.4 unit, caused great enlargement of the ovaries (Table 3) which were cystic; small corpora lutea were present. The vaginal epithelium consisted of stratified and keratinized epithelium with small areas of mucus cells. The adrenals were decreased in weight, and the pituitary glands were slightly enlarged (Table 3).

Following the cessation of gonadotrophin (0.4 unit) administered for 150 days the ovaries remained large and cystic and free from neoplasms during observation for an additional 5 months. Histological study of the vagina by smears and sections revealed that the rats were in a state of prolonged, usually persistent, estrus.

The mammary glands of rats fed 3-MC, 10 mg, and injected with equine gonadotrophin, 0.4 unit for 150 days were enlarged, orange in color and arborized. Intense epithelial proliferation was present consisting of luminated and solid tubules; many ducts and acini contained milk. Dilated milk-filled pools, 1 to 8 millimeters in diameter, were seen in the mammary glands 5 months after injections were discontinued.

Influence of 9α-Bromo-11-ketoprogesterone Administered Separately or Concurrently with Estradiol-17β.—Small doses of estradiol-17β (0.01 or 0.1μg) exerted no influence upon the incidence or on the time of appearance of mammary cancer in intact rats fed 3-MC (Table 4); vaginal smears of these rats showed no abnormality of the estrus cycle. Estradiol-17β, 1μg, also had no effect on mammary carcinogenesis; most of these rats were in a state of prolonged estrus.

The daily injection of estradiol-17β, 10μg, in conjection with 3-MC feeding, reduced the incidence of mammary cancer or delayed significantly its appearance (Table 4). These rats were in constant estrus while receiving estradiol-17β and there was hyperplasia of the mammary tree which contained milk.

BKP, 0.25 or 0.5 mg, administered in conjunction with 3-MC feeding accelerated the appearance of mammary cancer (Table 4). Smaller amounts of BKP, 0.05 or 0.1 mg, did not modify the incidence of mammary cancer.

In a series of experiments, 3-MC, 10 mg, was fed to groups of rats injected for 50 days with a fixed dose of BKP, 0.25 mg, and varying amounts of estradiol-17β; the dosage of both steroids remained fixed for the individual groups through the experiments. The injections of estradiol-17β, 0.1μg, together with BKP, 0.25 mg,

TABLE 4

INHIBITORY EFFECTS OF ESTRADIOL-17β ADMINISTERED CONCURRENTLY WITH 9α-BROMO-11-KETOPROGESTERONE

Steroids were injected, and 3-methylcholanthrene, 10 mg, was fed, 6 days each week for 50 days to intact female rats. Controls received only 3-methylcholanthrene.

Steroids; Dose	No. Rats	Rats with Cancer	Time of Appearance of Mammary Cancer, Days — Range	Median	Mean
		(*a*) Estradiol-17β			
None; controls	11	11	44–73	53	55.0 ± 8.8
Estradiol-17β, 0.01μg	9	9	42–62	55	52.9 ± 7.1
Estradiol-17β, 0.1μg	9	9	39–55	45	46.0 ± 5.4
Estradiol-17β, 1μg	9	9	41–81	58	59.4 ± 12.7
Estradiol-17β, 10μg	9	7	77–167	115	125 ± 33.9
		(*b*) 9α-Bromo-11-Ketoprogesterone (BKP)			
None; Controls	15	15	29–88	55	55.1 ± 12.2
BKP, 0.05 mg	10	10	29–57	49	48.7 ± 7.6
BKP, 0.1 mg	10	10	43–62	48	51.3 ± 7.1
BKP, 0.25 mg	10	10	38–49	42	43.2 ± 3.3
BKP, 0.5 mg	10	10	32–52	40	41.8 ± 5.6
		(*c*) Estradiol-17β plus BKP			
None; controls Estradiol-17β	11	11	44–73	53	55.0 ± 8.8
Estradiol-17β 0.01μg plus BKP 0.25 mg	9	9	38–67	45	49.3 ± 10.3
Estradiol-17β, 0.1μg plus BKP 0.25 mg	9	9	48–109	55	61.6 ± 18.7
Estradiol-17β, 1μg plus BKP 0.25 mg	9	6	51–161	105	108.0 ± 44.8
Estradiol-17β, 10μg plus BKP 0.25 mg	9	7	57–168	128	118.7 ± 39.6

±, Standard deviation.

retarded slightly the development of mammary cancer. Groups receiving estradiol-17β, 1μg or 10μg, together with BKP, had a significant decrease in incidence of mammary cancer, and a retardation of the time of appearance in those rats in which mammary cancer appeared.

The endocrine effects of BKP, 0.25 mg, injected daily for 50 days were studied in intact rats fed concurrently with 3-MC. The pregnane derivative caused an increase in body weight (Table 3) and a decrease of weight of the adrenals, ovaries, and uterus. Mucification was present in the vagina. There was a considerable progestational proliferation of the mammary glands, consisting of hyperplasia of small tubules, most of them having no lumen; milk was present in trace amounts. The effects were similar to those obtaining in rats treated similarly with large doses of progesterone (Table 3). There was no evidence that 3-MC had modified the effects of the pregnane derivatives except for the induction of mammary cancer. The concurrent administration of estradiol-17β, 1μg or 10μg, with BKP, 0.25 mg, increased profoundly the gestational effects in the vaginal epithelium and mammary glands; the mammary epithelium consisted of solid tubules and mere traces of milk were present.

Inhibitory Effect of Dihydrotestosterone.—Earlier[2] it was shown that dihydrotestosterone, 1 mg, considerably delayed the development of mammary cancer in rats fed concurrently with 7, 12-dimethylbenz-α-anthracene, 1 mg. In the present experiments, dihydrotestosterone, 3 mg, was injected for 50 days in 10 rats fed 3-MC, 10 mg, for the same period; 7 rats developed mammary cancer in 149 ± 43.9 days and 3 rats had no cancer after 10 months. The mammary glands at 50 days

when the injections were concluded were firm, greatly enlarged, and orange in color. On microscopic examination it was found that the mammary glands consisted essentially of a large number of luminated mammary tubules containing milk.

Discussion.—Confirming earlier work,[2] the intragastric administration of 3-MC, 10 mg, to intact, adult Sprague-Dawley rats beginning at age 50 days, either once or 6 times a week for 7 weeks led invariably to mammary carcinoma within 3 months. Hence, decreases in percentage of incidence and delay in time of appearance of the cancer have considerable significance.

Since many of the mammary carcinomas evoked by the rapid technique are hormone-dependent, it is clear that only hormone-independent tumors can emerge under adverse endocrine condition, e.g., the hormone-poor environment obtaining in hypophysectomized rats.

Earlier, Shay *et al.*[7] found that the concurrent administration of estradiol-17β or testosterone with small doses of 3-MC administered by gastric instillation reduced the incidence of breast cancer. In rats treated with physiologically effective doses of either of these steroids, there was increased proliferation of the mammary tree and lactation was present. From the present experiments, it would appear that in intact rats treated with 3-MC, estradiol-17β, in small doses, exerts no influence on the mammary carcinogenesis and in effective larger doses, acts as a blocking agent; estradiol-17β did not promote the development of mammary cancer.

New observations in the present work are (1) the absence of mammary cancer in hypophysectomized rats and (2) the reduced incidence of cancer of the breast in rats treated with compounds which caused a high degree of mammary proliferation.

Hypophysectomy is known[13] to inhibit effectively the formation of hepatomas in rats fed 3′-methyl-4-dimethylaminoazobenzene. In view of this observation plus the profound atrophy of the mammary tubules induced by hypophysectomy in the rodent, the complete failure of 3-MC to evoke mammary cancer in rats deprived of the pituitary was not unexpected.

Highly proliferative states of the non-neoplastic mammary glands of the rats are of 2 sorts, gestational and lactational. In the gestational proliferation, most of the mammary tubules have no lumina and secretion in the ducts is minimal; this hyperplasia occurs at Day 15 of normal pregnancy. In lactational proliferation, the mammary tubules are luminated and contain milk; proliferation of this type was present in rats treated with amounts of estradiol-17β or testosterone sufficient to inhibit or delay mammary carcinogenesis.

Large doses of equine gonadotrophin (e.g., 0.4 unit for 150 days) evoked intense ovarian activity resulting in highly proliferated mammary glands of lactational type. In these animals, the ovaries remained large and cystic for at least 5 months after gonadotrophin injections had been discontinued and the mammary glands contained cystic pools of milk visible in the gross. In such animals treated with 3-MC for the first 50 days of the experiment, the incidence of mammary cancer was reduced to approximately 25 per cent of the controls. It is noteworthy that every rat receiving this dosage of equine gonadotrophin in the present experiments for 150 days developed sarcoma at the injection site. Earlier, Bielschowsky (14) observed that established mammary cancer induced by feeding 2-AAF frequently regressed, sometimes profoundly, during lactation. Haddow[15] and Foulds[16] have observed regression of spontaneous mammary cancer during lactation in mice.

The most interesting finding in the present experiments, in which rats were fed 3-MC, is the inhibitory effect of estradiol-17β on the development of mammary cancer in animals receiving BKP. BKP is the most effective agent known at the present time in accelerating the development of mammary cancer in intact rats and it induces pronounced gestational proliferation in their breasts. But estradiol-17β, 1 to 10μg administered supplementary to BKP, 0,25 mg, blocked the development of mammary cancer in some rats and considerably delayed its development in the remainder while enhancing gestational growth profoundly. The most reasonable explanation for the inhibitory effects of steroids on mammary carcinogenesis is that their presence in the receptor sites of the mammary tubules blocks the concentration of 3-MC in the mammary tree and, hence, the rapid development of mammary cancer.

Summary.—The physiologic state of the mammary tree, dependent on endocrine effects, is of critical significance in rats fed 3-methylcholanthrene. Every normal female control rat developed mammary cancer before 92 days. Cancer of the breast did not form in the atrophic mammary glands of hypophysectomized rats. But the incidence of breast cancer also was significantly reduced in rats with profound mammary hyperplasia, of gestational or lactational type, induced in several ways; a highly effective blocking procedure was the administration of amounts of equine gonadotrophin sufficient to produce intense and long continuing stimulation of the ovaries and of the mammary gland with persistent lactation.

9-α-Bromo-11-ketoprogesterone is the most effective agent as yet discovered in accelerating the development of mammary carcinoma in intact rats fed 3-MC. This BKP-induced acceleration of cancer development was abolished by the concurrent administration of estradiol-17β; the incidence of mammary cancer was significantly reduced or the appearance of tumors was delayed while at the same time the gestational effects of BKP were profoundly enhanced by the phenolic estrogens.

* This work was supported by grants from the American Cancer Society, The Jane Coffin Childs Memorial Fund for Medical Research, and the U. S. Public Health Service.

[1] The abbreviations used in this paper are: 3-MC, 3-methylcholanthrene; 2-AAF, N2-fluorenylacetamide; BKP, 9α-bromo-11-ketoprogesterone.

[2] Huggins, C., G. Briziarelli, and H. Sutton, Jr., *J. Exp. Med.*, **109**, 25 (1959).

[3] Maisin, J., and M.-L. Coolen, *Compt. rend. Soc. biol.*, **123**, 159 (1936).

[4] Bielschowsky, F., *Brit. J. Exp. Path.*, **25**, 1 (1944).

[5] Geyer, R. P., J. E. Bryant, V. R. Bleisch, E. M. Peirce, and F. J. Stare, *Cancer Research*, **13**, 503 (1953).

[6] Shay, H., E. A. Aegerter, M. Gruenstein, and S. A. Komarov, *J. Nat. Cancer Inst.*, **10**, 255 (1949).

[7] Shay, H., M. Gruenstein, and C. Harris, *Proc. Am. Assoc. for Cancer Research* **2**, 146 (1956).

[8] Cantarow, A., J. Stasney, and K. E. Paschkis, *Cancer Research*, **8**, 412 (1948).

[9] Rockland Rat Diet, A. E. Staley Co., Chicago, Ill.

[10] Equine gonadotrophin (GONADOGEN) was generously provided by The Upjohn Company Kalamazoo, Michigan.

[11] Cartland, G. F., and J. W. Nelson, *Am. J. Physiol.*, **122**, 201 (1938).

[12] Huggins, C., and E. V. Jensen, *J. Exp. Med.*, **100**, 335 (1955).

[13] Griffin, A. C., A. P. Rinfret, and A. F. Corsiglia, *Cancer Research*, **13**, 77 (1953).

[14] Bielschowsky, F., *Brit. Med. Bull.* **4**, 382 (1947).

[15] Haddow, A., *J. Path. Bact.*, **47**, 553 (1938).

[16] Foulds, L., *Brit. J. Cancer*, **3**, 345 (1949).

MICROSCOPIC TESTS OF SYMMETRY PRINCIPLES

By G. Feinberg and M. Goldhaber

BROOKHAVEN NATIONAL LABORATORY*

Read before the Academy, April 29, 1959

In the analysis of the phenomena involved in the interaction of elementary particles, use is often made of the implications of "absolute" conservation laws, such as the conservation of charge, the conservation of baryons, and the conservation of leptons. It is known from the general principles of quantum mechanics that the existence of such conservation laws is connected with certain symmetries which are believed to be possessed by the equations describing the behavior of the elementary particles. Some of the conservation laws are familiar from classical physics, because the conserved quantity which appears in them can be directly measured by macroscopic experiments (e.g., the charge). In other cases, a conservation law appears only indirectly in classical experiments, and the evidence for it derives from experiments involving elementary particles.

While no apparent contradictions have arisen from the application of the absolute conservation laws, and the symmetries with which they are connected may appear quite reasonable on theoretical grounds, we should remember that in the history of physics it has been demonstrated many times, most recently in the discovery of the non-conservation of parity in weak interactions, that the basic principles of theoretical physics cannot be accepted *a priori*, no matter how convincing they may seem, but rather must be justified on the basis of relevant experiments. The acceptance of this empiricist view of symmetry principles has an obvious consequence. One would like to know what the specific experimental evidence for a particular symmetry is, both in the qualitative sense of knowing which experiments bear on it, and in the quantitative sense of knowing what the limits are of possible violations of the particular symmetry. It is often the case that we believe in certain conservation laws, not so much because they have been explicitly tested to a considerable degree, but because they are implicitly used in many experiments without leading to difficulties. It is therefore desirable to replace, wherever possible, implicit and qualitative statements concerning the validity of conservation laws by explicit tests leading to quantitative limits.

Of course, it has been recognized for a long time that the final justification of symmetry principles, as of all other physical laws, is the experimental evidence for them. Some of the experiments we shall cite in this paper were done some time ago, and it is possible that still others, unknown to us, have been done which have an important bearing on this question. Nevertheless, it appears worth while to collect in this preliminary form some samples of the experiments which put rather stringent limits on possible violations of conservation laws and related symmetry principles. We shall concentrate on some of those symmetries which are believed to be universally valid, rather than restricted to a particular class of interactions. We do this because the evidence about other types of symmetry such as parity, charge conjugation, charge independence, etc., has been the subject of much recent analysis[1] and we have nothing new to add to it. In the second section we shall

discuss the evidence for the absolute conservation laws connected with gauge transformations, i.e., the conservation of charge, of baryons and leptons. In the third section we shall discuss the evidence for the equality of the charges of the various charged elementary particles, a symmetry which corresponds to no known conservation law, but which is nevertheless rather striking, since, as we shall see, the assumption of charge conservation and the existence of the known reactions among the elementary particles do not by themselves imply all of the charge equalities which appear to exist. In the fourth section we discuss some of the evidence for the conservation of angular momentum, which is associated with symmetry under space rotations. It may be useful to state here that we do not know of any evidence against any of the symmetries which we discuss. However, the rather extreme experimental limits which can be placed on violations of the conservation of charge and of baryons indicate that any attempt to give up these laws, even in a restricted class of interactions, must be made with great care.

The Absolute Conservation Laws for Additive Quantum Numbers.—By an additive quantum number one means a quantity which can be assigned to each particle, such that its value for a system of particles is the algebraic sum of its values for the separate particles and is independent of the state of the system. There are examples of additive quantum numbers, such as "strangeness" which are conserved in some classes of reactions but not in others. When such a quantity is conserved in all reactions, one speaks of an "absolute conservation law" for the quantity. It can be shown that the existence of absolute conservation laws is associated in a field theory with the invariance of the total Lagrangian under a gauge transformation in which the field operators for all of the particles are multiplied by phase factors of the form

$$\psi_N \rightarrow e^{iN\lambda}\psi_N \tag{1}$$

where ψ_N is the field operator for a particle with the value N for the quantum number in question and λ is an arbitrary number which is the same for all the particles. There are three conservation laws which are believed to be absolute in the present theory of elementary particles. These are the conservation of electric charge, the conservation of baryons, and the conservation of leptons. We shall now discuss some experiments which give limits on possible violations of these conservation laws.

Conservation of electric charge: For the known elementary particles, the charges take on only a small number of values which appear to be integral multiples of the electron charge. It is therefore not possible to have transitions among these particles in which the charge changes by some fraction of this unit. If there were interactions which do not conserve charge, these would produce transitions from an initial state containing particles whose total charge is some integral multiple of the unit charge to a final state containing particles with a total charge which is a different multiple of the unit charge. A measure of charge nonconservation will then be the branching ratio for such transitions compared to other transitions from the same state, in which the total charge is unchanged. The question of why charge is quantized is a separate question with an as yet unknown answer, which we return to, but do not solve, in the next section. The situation here may be compared with that for angular momentum, where it follows from the quantum

conditions on angular momenta, that even if angular momentum were not conserved, it could only change by multiples of $\hbar/2$.

In discussing conservation laws, it is useful to consider, whenever possible, experimental effects which depend only on the violation of the conservation law in question, leaving the specific mode of violation as arbitrary as possible. Since we are looking for effects which are much smaller than those with which we are familiar, it would not be wise to assume that the principles governing known interactions, such as the assumption that only a small number of particles are involved in each elementary process, are valid. One may even consider the possibility that several of the conservation laws may be violated simultaneously. Some of the experiments we consider are more or less independent of the mode of violation of the conservation laws, while others are not. In the discussion of the significance of some of the experiments we shall consider particular modes of possible violations of the conservation laws. These are only used for illustration.

For some possible violations of conservation laws, limits have been deduced from macroscopic considerations, such as the apparent "electric neutrality" of macroscopic bodies. However, to make these arguments precise involves in some cases a detailed analysis of compensating phenomena. It is therefore desirable to set limits for possible violations in elementary processes, though they are sometimes not so stringent as those suggested by macroscopic arguments.

A failure of charge conservation would allow processes to occur which are otherwise forbidden. Clearly, a particularly good place to look for such violations is to examine systems which cannot make any transitions at all if the conservation laws are satisfied, but could if charge were not conserved. An example of such a system is an electron, either free, or bound in an atom which is in its ground state. The electron is the state of lowest rest mass with an electronic charge. However, there are states of lower rest mass into which the electron could decay if charge were not conserved, such as a state with one neutrino and one photon, or two neutrinos and an antineutrino. Therefore, the "lifetime" of a free electron, or an electron in the ground state of an atom, is a test of charge conservation. Strictly speaking, electron decay might occur into unknown charged particles with mass smaller than the electron mass, without violating the conservation of charge. The existence of such particles is presumably ruled out by the fact that they have not been produced in pairs by photons.[2]

The possibility of electron decay was recently investigated at this Laboratory.[3] One of the consequences of the decay of atomic electrons would be the formation of K-electron "holes," which could be well detected in a NaI scintillation counter by the subsequent emission of an X ray. A search for the iodine K X rays leads to a lower limit for the electron "life-time" $\tau_e > 10^{17}$ years.

It should be noted that this experiment is sensitive to the decay of electrons by any mode since all that is involved is the disappearance of an inner electron. The lower limit for the partial lifetime for some specific modes of decay can be made larger; e.g., for the process $e \rightarrow \nu + \gamma$, which would lead to energetic γ-rays, more easily distinguished from background, it is found to be $>10^{19}$ years.

Another test of charge-conservation comes from the stability of certain nuclei. Consider two nuclei whose atomic number and mass number are given by (Z, A) and $(Z + 1, A)$. Let us distinguish between nuclear masses (m) and atomic masses

(M). Suppose that the nuclear masses of the two nuclei are such that $m_{Z+1} < m_Z < m_{Z+1} + m_e$, where m_e is an electron mass. In this case, the nucleus (Z, A) cannot decay by beta emission into $(Z + 1, A)$ because there is not enough energy available to create an electron. However, if charge-nonconserving decays were allowed, the decay could occur via processes which have zero energy threshold, such as neutrino pair emission or photon emission. The latter process might be directly detected by conventional methods. Alternatively, one may make use of the fact that the nucleus $(Z + 1, A)$ is itself unstable against K capture, since the masses of the atoms are related by $M_Z < M_{Z+1}$. It is therefore possible to search for the decay of (Z, A) by detecting the radioactive product nucleus, though, to our knowledge, this has not yet been done. (An example of a nuclear pair which fulfills these energy conditions is ${}_{19}K^{41}$ and its isobar ${}_{20}Ca^{41}$.)†

It is of some interest to estimate the maximum strength of interactions consistent with the limits given above for the electron decay. Such estimates must necessarily depend on the models assumed for the violation of charge conservation. However, we believe that they will give some indication of the orders of magnitude involved.

Suppose that there is an interaction producing the decay $e \rightarrow \nu + \gamma$. This could, for example, be represented by an effective matrix element

$$M = \frac{\lambda_c}{m} \bar{\psi}_\nu \sigma_{\alpha\beta} \psi_e F_{\alpha\beta}, \tag{2}$$

where ψ_e is the electron wave function, ψ_ν the neutrino wave function, $F_{\alpha\beta}$ is the electromagnetic field tensor and $\sigma_{\alpha\beta}$ is the spin tensor. Here λ_c is a dimensionless number measuring the interaction strength, while m is a mass which we shall take to be the mass of the electron. The lifetime for the free electron calculated with this matrix element is

$$\tau_e = \frac{2\pi}{\lambda_c^2} \frac{\hbar}{m_e c^2} = \frac{7 \times 10^{-21}}{\lambda_c^2} \text{ sec.} \tag{3}$$

Using $\tau_e > 10^{19}$ years $= 3 \times 10^{26}$ sec., we find

$$\lambda_c^2 < 2 \times 10^{-47}. \tag{4}$$

This number may be taken as an indication of the weakness of any possible decay interactions involving electrons and not conserving charge. The number may be compared with the dimensionless Fermi coupling constant g which we here define in terms of the conventional Fermi coupling constant G_F, by

$$G_F = \frac{g}{m_e^2},$$

again using m_e as the standard mass for comparison purposes, giving

$$g^2 = 10^{-23}.$$

It may be seen that any charge-nonconserving interactions of the type discussed are, if they exist, many orders of magnitude weaker than the "weak interactions." The limit on the dimensionless coupling constant for decays like $e \rightarrow 2\nu + \bar{\nu}$, which follows from the above experiments is similar. One can conclude from a correspond-

ing calculation that $g_e^2 < 10^{-42}$, where the interaction leading to $e \rightarrow 2\nu + \bar{\nu}$ is taken to be:

$$\frac{g_e}{m_e^2} \bar{\psi}_\nu \psi_e \bar{\psi}_\nu \psi_\nu. \tag{5}$$

It is also desirable to investigate possible interactions which violate charge conservation by several units at once, i.e., "modulo n." By this is meant that, while processes which change the charge by 1, 2, ... $n - 1$ units at once are forbidden, these interactions permit processes which change it by n, $2n$, $3n$, etc. units. Selection rules of this kind occur when one deals with multiplicative quantum numbers rather than additive ones. Such theories are proposed from time to time. Evidence against some interactions of this type is provided by the stability of many-electron atoms, as in the experiment quoted above. For instance, if it were possible for several electrons to annihilate at a point, with the emission of neutral particles, one would expect this to happen in elements such as iodine, where the overlap of the electron wave functions is substantial. This would again lead to the "disappearance" of electrons in the inner shells and the subsequent emission of X radiation. A comparison of the rate of this type of annihilation with the lower limit on the lifetime of these atoms gives upper limits for the coupling strengths for interactions like $4e \rightarrow 4\nu$ which are comparable to the limit quoted for g_e above.

Conservation of Baryon number: We next consider the conservation of baryons, a law first suggested by Wigner[4,5] in view of the apparent stability of the proton, and later found to be useful in discussing the decays of hyperons, which end up in a proton or a bound neutron. This law states that in all elementary particle processes there exists a conserved additive quantum number which is chosen as $+1$ for nucleons and hyperons, -1 for their antiparticles and 0 for all other known particles. Since the proton is the lightest known particle with the value $+1$ for this quantum number, the conservation of baryons guarantees the stability of the free proton against decay into known particles. It would appear that the "lifetime" of the proton, either free, or bound in nuclei which are stable against the ordinary decays, is a sensitive test of the existence of interactions which do not conserve baryons.[6]

An experiment which places a limit on the proton lifetime for certain decay modes has been done by Reines, Cowan, and Goldhaber,[7] who looked for possible proton decays in a large hydrogeneous scintillation counter (C_7H_8). The limit obtained for the free proton lifetime is $\tau_p > 10^{21}$ years and for bound protons $> 4 \times 10^{23}$ years.[8]

Again, an experiment which is independent of the method of decay is useful because it involves the fewest extensions of present ideas into realms where they may not apply. An experiment of this kind is provided by the stability of certain heavy nuclei against spontaneous fission.[7]

From the absence of spontaneous fission in Th^{232}, Flerov *et al.*[9] find a limit $> 2 \times 10^{23}$ years for bound nucleons decaying by any mode.

An analysis of the upper limit of the strength of baryon nonconserving interactions, based on a hypothetical process such as $P \rightarrow e^+ + \pi^0$, can be done, similar to the one carried out for charge nonconservation. If we take the matrix element for this process as

$$M = \lambda_\beta \bar{\psi}_e \psi_p \phi_{\pi^0}, \tag{6}$$

where ψ_p is the proton wave function, ϕ_{π^0} is the π^0 wave function and λ_β is a dimensionless constant. We find that

$$\tau_p \simeq \frac{\hbar}{m_p c^2} \frac{1}{\lambda_\beta^2} < 10^{29} \text{ sec.}$$

or

$$\lambda_\beta^2 < 10^{-53}. \tag{7}$$

Similar values are obtained for the dimensionless coupling constants for processes[10] like

$$P \rightarrow 2e^+ + e^-,$$

provided that all quantities with the dimension of a mass are taken to be about m_p.

The results for nucleons bound in nuclei also restrict the possibility of interactions in which nucleon number is changed "modulo n." In heavy nuclei it is to be expected that the overlap of the wave functions for n nucleons is appreciable $\left[\sim \left(\frac{r_{nuc.}}{r_0}\right)^{3(n-1)}\right]$. Here $r_{nuc.}$ is the "mean extension" of the nucleon, while r_0 is the "average separation of nucleons" in a nucleus, or 1.2×10^{-13} cm. One sees that interactions involving several nucleons at a point will not be reduced by many orders of magnitude, unless the matrix elements have some special space or spin dependence. Like the processes discussed for single nucleon decay, processes where many nucleons are destroyed simultaneously would in general give charged decay products, or lead to a product nucleus unstable against fission. We therefore find, for example, that the square of the dimensionless coupling constant for interactions which are "baryon nonconserving modulo 2" must be $<10^{-49}$.

Conservation of leptons: The third absolute conservation law which is generally believed is the conservation of leptons. The evidence for this law in the usual weak interactions has been the subject of much analysis since the discovery that neutrinos and antineutrinos differ in their helicity. We will not review these discussions, except to note that the conclusion is that any lepton nonconserving terms in processes such as β decay, π-μ decay and K-μ decay are probably <10 per cent of the lepton conserving terms.[11]

As far as we know now, the leptons do not possess any interactions other than minimal electromagnetic interactions,[12] or weak interactions involving neutrinos. However, the experimental limits on the strength of other lepton-involving interactions are not very severe. For instance, from the fact that the decay $\pi^0 \rightarrow e^+ + e^-$ is no more than 10^{-4} of all π^0 decays,[13] it can be concluded that an interaction

$$\lambda_\pi \bar{\psi}_e \psi_e \phi_{\pi^0} \tag{8}$$

must satisy $\lambda_\pi^2 < 10^{-10}$.

From the lack of existence of the decay $\mu \rightarrow e + \gamma$($10^{-6}$ of ordinary μ decays[14]) it follows that the dimensionless coupling constant g_μ for an interaction which could produce this decay,

$$\frac{g_\mu}{m_e}\bar{\psi}_e\sigma_{\alpha\beta}\psi_\mu F_{\alpha\beta}, \tag{9}$$

where ψ_μ is the μ meson wave function, satisfies $g_\mu^2 < 10^{-27}$.

It is also possible to obtain severe limits on some conceivable leptonic interactions which do not involve neutrinos and which do not conserve the number of leptons. As a hypothetical interaction of this kind, consider a coupling which changes $2\pi^-$ into $2\ e^-$,

$$M = g_L/m_e(\phi_{\pi^-})^2\bar{\psi}_e{}^c\psi_e. \tag{10}$$

Here $\psi_e{}^c$ is the charge conjugate of the electron wave function. This interaction, when combined with the strong π-nucleon interaction, would lead to processes such as

$$(Z,A) \rightarrow (Z + 2, A) + 2e^-,$$

a double β decay process. Note that this process would now occur in the first order of this assumed lepton nonconserving interaction, whereas double beta decay via the often discussed lepton nonconserving interactions involving neutrinos is a second order process. An order-of-magnitude estimate of the rate for double β decay via the new interaction is given by the rate of single β decay via the Fermi coupling, multiplied by $(g_L/g)^2$. In the well investigated case of Ca^{48} one finds a limit for the lifetime for double β decay $> 7 \times 10^{18}$ years,[15] whereas, for a corresponding energy release, the lifetime for single β decay via the Fermi interaction would be $\sim$1 sec. From this it may be concluded that the coupling strength for interactions like the above which change the lepton number by 2, must be 10^{13} times weaker than the Fermi coupling, or

$$g_L{}^2 \lesssim 10^{-49}. \tag{11}$$

It is also of some interest to consider a more general case of possible interactions in which leptons are not conserved modulo n (>2, even). If such interactions exist, it is expected that decays of the type

$$(Z, A) \rightarrow (Z + n, A) + ne^- \qquad \text{or}$$

$$(Z, A) \rightarrow (Z + n - m, A) + (n - m)e^- + m\nu$$

could occur when they are energetically possible. We are not aware of any search for such decay modes.

Charge Equalities of the Elementary Particles.—We have seen in the last section that the evidence for the three absolute conservation laws is quite strong. To start with, we will therefore assume in this section that these laws are satisfied rigorously. The existence of these three conservation laws, rather than just the conservation of charge, is relevant to the problem which has often been raised of why the charges of all the elementary particles are $\pm Q_e$ or zero. To demonstrate the connection between the conservation laws and the charge equalities, we note that the conservation laws state that there are three independent sets of numbers Q_i, B_i, L_i, the charges, baryon numbers, and lepton numbers of the particles, which satisfy

$$\sum_i Q_i = \text{constant in time},$$
$$\sum_i B_i = \text{constant in time}, \tag{12}$$
$$\sum_i L_i = \text{constant in time}.$$

It is clear, however, that if this is the case, then also any independent set of linear combinations of these numbers, such as

$$\begin{aligned} Q_i' &= a_1 Q_i + b_1 B_i + c_1 L_i \\ B_i' &= a_2 Q_i + b_2 B_i + c_2 L_i \\ C_i' &= a_3 Q_i + b_3 B_i + c_3 L_i \end{aligned} \tag{13}$$

will also satisfy the relations (12). Furthermore, since the a, b, c are arbitrary, except that the determinant formed by them must be nonzero, it follows that if the ratio of the Q_i are integers, the ratio of the Q_i' will not necessarily be so. Thus, because of the conservation of baryons and leptons, the application of conservation of charge to the known reactions involving elementary particles does not, of itself, determine the ratio of the charges of all the elementary particles. Instead, there will be two ratios which are left arbitrary by all reactions. For instance, the apparent absence of $p \to e^+ + \pi^0$ and of similar processes leaves the relative charge of p and e undetermined. Even if one assignment of charges, Q_i, leads to a conservation law with all charges integral, another conserved assignment would be obtained by adding to the charge of each particle a fixed multiple of its baryon number or lepton number. The same holds for the assignments of baryon or lepton number, i.e., one could for each particle add a fixed multiple of the other two additive quantum numbers and obtain a new assignment of baryon or lepton number which is conserved. This would leave unchanged the quantum numbers only of those particles for which all conserved additive quantum numbers have the value zero (e.g., the π^0).

These results are summarized, for the case of charge, in Table 1, which lists in four columns those particles whose charges are required to be equal because of existing reactions. Also listed, in brackets, is a reaction implying this result.

TABLE 1

$Q = 0$	$Q = Q_p$	$Q = Q_n$	$Q = Q_{e^-}$
$\gamma(p + p \to p + p + \gamma)$	p	n	e^-
$\pi^0(p + p \to p + p + \pi^0)$	$\Sigma^+(\Sigma^+ \to p + \pi^0)$	$\Lambda^0(\Lambda^0 \to n + \pi^0)$	$\mu^-(\mu^- \to e^- + \nu + \bar{\nu})$
$\theta_1^0(\theta_1^0 \to 2\pi^0)$		$\Sigma^0(\Sigma^0 \to \Lambda^0 + \gamma)$	
$\theta_2^0(\theta_2^0 \to \pi^+ + \pi^- + \pi^0)$		$\Xi^0(\Xi^0 \to \Lambda^0 + \pi^0)$	

We also have the following relations:

$$Q_{\text{particle}} = -Q_{\text{antiparticle}} \text{ (particle + antiparticle} \to 2\gamma)$$
$$\begin{aligned} Q_{\pi^+} &= Q_p - Q_n \ (p \to n + \pi^+) \\ Q_{K^+} &= Q_{\pi^+} \ (K^+ \to \pi^+ + \pi^0) \\ Q_{\Sigma^-} &= 2Q_n - Q_p \ (\Sigma^- + d \to \Lambda^0 + 2n) \\ Q_{\Xi^-} &= 2Q_n - Q_p \ (\Xi^- \to \Lambda^0 + \pi^-) \\ Q_\nu &= Q_p + Q_{e^-} - Q_n \ (n \to p + e^- + \nu^-) \end{aligned}$$

These equations exhaust the relations among the charges of elementary particles implied by the conservation of charge in the known reactions.[16] It can be seen that after fixing the scale of charge by measuring in terms of Q_{e^-}, there are still two charges left undetermined by these relations, which we can choose as Q_n and Q_p.

It is important to realize that the charges of particles are in fact measured, not by an appeal to conservation laws, but by placing the particles in external electromagnetic fields and observing their behavior. For the baryon number and lepton number, however, there is no evidence that fields which interact specifically with the "baryonic charge" or "leptonic charge" actually exist,[11] and so it is doubtful that any physical significance can be attributed to any one of the possible conserved assignments of these quantities in preference to others, at least at present. Thus, only for the electric charge is it possible to decide experimentally which of all the possible conserved assignments of Q_f is correct. The results of various experiments are that the possible freedom in the relative electric charges of the particles is practically removed. Consider first the charge of the neutron. A possible small charge for the neutron could be detected by the deflection of a neutron beam by a homogeneous electric field. Such an experiment has been carried out by Shapiro and Estulin,[18] who find an upper limit for the charge of the neutron $\leq 6 \times 10^{-12}|Q_{e^-}|$. A similar experiment has been done by Hughes,[19] on the *CsI* molecule, who finds

$$|Q_{CsI}| < 4 \times 10^{-13} |Q_e| \tag{14}$$

Using

$$Q_{CsI} = 108\ (Q_p + Q_e) + 152\ Q_n,$$

and assuming that there is no accidental cancellation between the two terms, one can conclude that

$$\begin{aligned} |Q_p + Q_{e^-}| &< 4 \times 10^{-15} |Q_{e^-}| \\ |Q_n| &< 3 \times 10^{-15} |Q_{e^-}|. \end{aligned} \tag{15}$$

Alternatively, one can combine this result with the previous result on the neutron charge and conclude that

$$|Q_p + Q_{e^-}| < 9 \times 10^{-12} |Q_{e^-}|, \tag{16}$$

whether or not there is cancellation. The assumption of charge conservation then implies that the charges of all the other elementary particles are equal to the values usually assigned, within limits which are comparable to equations (15) and (16).

These results make it seem reasonable that all the known particles have charge $\pm Q_{e^-}$ or 0. This would follow on the basis of our previous discussion, if the conservation of baryons and leptons were only approximate, and the conservation of charge were the only absolute conservation law, since there would then be only one possible assignment of charges. As mentioned above, if processes like $p \rightarrow e^+ + \pi^0$ were observed, then charge conservation would require $Q_p = Q_{e^+}$. Conversely, if it were found that the electric charges of the baryons were all slightly displaced from their usually accepted values, say by the common value ϵ, the conservation of baryons would follow from the conservation of charge, instead of being an independent physical principle.[20] Similar remarks hold for leptons. However, the results of the previous section give little encouragement to these "explanations." It there-

fore seems that the apparent equality of the charges of the elementary particles involves some new physical principle.

Conservation of Angular Momentum.—It is well known that the invariance of particle interactions under rotation of the coordinate system implies the existence of a conserved axial vector, the angular momentum of the system of particles. The consequences of this conservation law have been very useful in many fields of physics, such as atomic and nuclear spectra, angular correlations, etc. It is useful to enquire about a possible violation of this conservation law, which violation would presumably imply some lack of isotropy of space.

One consequence of the conservation of angular momentum is that the stationary states of every system are also eigenstates of the square of the angular momentum operator and one of its components, in the absence of accidental degeneracy. This allows the usual classification of nuclear or atomic states in terms of the eigenvalues of J^2, J_z, etc. A violation of the conservation of angular momentum in the interactions which bind nuclei and atoms would, among other things, lead to admixtures of several different total angular momenta into a given stationary state. If this occurs there will be, in transitions between these states, violations of the selection rules which are based on the usual assignments of angular momentum. This will occur whether or not angular momentum is conserved in the interactions which produce the transitions between the states. As an example, consider two states of a nucleus, or an atom, which have angular momentum zero. Then a transition between these two states with the emission of one photon is forbidden by the conservation of angular momentum, since the photon has spin 1. If, however, the nuclear force has a part which does not conserve angular momentum, then one would expect that these states could have some admixture of angular momentum one. In that case the emission of a photon can proceed either by electric or magnetic dipole radiation, even if no change is made in the electromagnetic interaction.[21]

An experiment to detect γ transitions between two nuclear states of "spin O^+" has been performed by Sunyar.[22] He searched for γ-rays in competition with the 700 Kev $O^+ \rightarrow O^+$ internal conversion electron transition in Ge^{72}. He found $\gamma/e^- < 1/1000$. Since the γ rate for a $1^\pm \rightarrow O^+$ transition would be of the order of 10^6 faster than the e^- rate for the $O^+ \rightarrow O^+$ transition, we can conclude that the amplitude for admixture of a spin 1 state is $< \sqrt{10^{-9}}$ (assuming normal matrix elements).

Other evidence for the nonmixing of angular momenta in stationary states comes from the approximate validity of the forbiddenness rules for estimating lifetimes of nuclear states against β or γ decay (or of atomic states against radiative transitions). The decay rate for "forbidden" transitions is generally reduced by a factor like $(ka)^{2n}$ where k is the energy release in natural units, a is some length characterizing the nucleus, and n is the order of forbiddenness. From an analysis of lifetimes, therefore, one can place limits on the admixture of states with angular momentum $J \pm \Delta J$, into a state whole "main" angular momentum is J, of the order

$$f \sim (ka)^{\Delta J}, \tag{17}$$

where f is the amplitude for admixture. Any appreciably greater admixture would make the lifetime much shorter than the forbiddenness rule would allow. In this way, one may conclude that, for a given state, limits on the admixing amplitudes

for angular momenta differing by several units must be small since ka may typically be $\sim 10^{-3}$. Clearly, this discussion is oversimplified in that the effects of other approximate selection rules are omitted. Some improvement in these estimates might be obtained by an analysis of angular correlations, or spectrum shapes, which are sometimes more sensitively dependent on spin changes than are lifetimes.

We would like to thank Dr. V. Hughes, Dr. N. Kroll, and Dr. H. Snyder for helpful discussions regarding several points.

* Under the auspices of the U. S. Atomic Energy Commission.

† *Note added in proof:* A pair of nuclei which are more amenable to an experimental test, and which differ by $\sim 0.8 m_e$ in their nuclear masses are Rb^{87} and Sr^{87m}. An attempt was made to detect the 390 kev γ-rays from Sr^{87m} (2.7 hours), extracted from a Rb sample. This yields a lower limit for the partial lifetime of Rb^{87} decaying by a charge-non-conserving process to Sr^{87m} of $\tau > 10^{16}$ years. (A. W. Sunyar and M. Goldhaber, unpublished).

[1] See Lee, T. D., and C. N. Yang, "Elementary Particles and Weak Interactions," BNL 443.

[2] Electrons bound in atoms, or free, could also "disappear" if instead other conservation laws, such as the conservation of energy, were violated. If one detects all electron "decays," as in the experiment described below, this places an upper limit on the number of electrons decaying by transitions which do not conserve charge.

[3] derMateosian, E., and M. Goldhaber (unpublished).

[4] Wigner, E. P., *Proc. Am. Phil. Soc.*, **93,** 521 (1949).

[5] Wigner, E. P., these PROCEEDINGS, **38,** 449 (1952).

[6] It should be recognized that decays of a proton involving a hitherto unknown particle would not necessarily violate the conservation of baryons, even if this particle had a quite small mass, provided that the new particle were itself stable, since in this case one would simply assign baryon number +1 to the new particle. On the other hand, the occurrence of processes like $n + n \rightarrow \gamma + \gamma$ would automatically involve violation of the conservation of baryons, since the photon cannot have a nonzero baryon number because it may be produced singly, as in $p + p \rightarrow p + p + \gamma$.

[7] Reines, F., C. L. Cowan, Jr., and M. Goldhaber, *Phys. Rev.*, **96,** 1157 (1954).

[8] Reines, F., C. L. Cowan, Jr., and H. W. Kruse, *Phys. Rev.*, **109,** 609 (1957).

[9] Flerov, G. N., D. S. Klochov, V. S. Skobkin, and V. V. Terentiev., *Soviet Physics, Doklady* **3,** 78 (1958).

[10] Y. Yamaguchi (private communication) recently has considered the possibility of the decay $p \rightarrow 2e^+ + e^-$, $2p \rightarrow 2\pi^+$ and other baryon nonconserving interactions and obtained similar limits on the coupling constants. We would like to thank Dr. Yamaguchi for sending us his preprint.

[11] See Goldhaber, M., "1958 Annual International Conference on High Energy Physics." (Geneva: CERN), p. 234.

[12] Minimal electromagnetic interactions are interactions of the electromagnetic field with the current of charged particles. See Pais, A., *Phys. Rev.*, **86,** 663 (1952); Gellmann, M., *Proc. Sixth Ann. High Energy Conference* (1956).

[13] Samios, N., and J. Steinberger, private communication.

[14] Davis, H. P., A. Roberts, and T. F. Zipf, *Phys. Rev. Lett.*, **2,** 211 (1959); D. Berley, J. Lee, and M. Bardon, *Phys. Rev. Lett.*, **2,** 357 (1959).

[15] Dobrokhotov, E. I., Ya Lazarenko, and S. Yu. Luk'yanov, private communication (see ref. 11).

[16] Since the K^0 and $\bar{K}^0$ mesons are linear combinations of θ_1^0 and θ_2^0 they must also be electrically neutral. This also follows from the production process $\pi^- + p \rightarrow \Lambda^0 + K^0$ and the above relations

[17] The possible existence of a "baryon gauge field" has been discussed by Lee, T. D., and C. N. Yang, *Phys. Rev.*, **98,** 1501 (1955), who point out that if it exists, it must interact extremely weakly. The existence of a "lepton gauge field" has also been considered by many physicists.

[18] Shapiro, I. S., and I. V. Estulin, *Soviet Physics. JETP*, **3,** 626 (1957).

[19] Hughes, V., *Phys. Rev.*, **105**, 170 (1957). This paper also discusses several macroscopic experiments which give considerably smaller limits for the charges of molecules.

[20] If the charge of the proton were $(1 + \epsilon)\,|Q_e-|$ then, e.g., the decay $p^+ \rightarrow e^+ + \pi^0$ would be forbidden. However the stability of an arbitrarily large number of protons would not be guaranteed by this alone, unless ϵ is irrational. Formally, this follows from the fact that then a subgroup of the charge gauge transformations (1) forms a dense subgroup of the baryon gauge transformations, and this implies baryon conservation.

[21] Of course, if angular momentum were not conserved, there might be additional electromagnetic interactions which would produce $O^+ \rightarrow O^+$ γ-transitions. However, the effect of such new interactions will depend critically on the type of interaction assumed and so we do not discuss them here.

[22] Sunyar, A. W., private communication.

---

*MINIATURIZATION OF THE ELECTROMAGNETIC BLOOD FLOW METER AND ITS USE FOR THE RECORDING OF CIRCULATORY RESPONSES OF CONSCIOUS ANIMALS TO SENSORY STIMULI**

By Alexander Kolin and Raymond T. Kado

DEPARTMENT OF BIOPHYSICS, UNIVERSITY OF CALIFORNIA AT LOS ANGELES

Communicated by James Franck, May 29, 1959

Introduction.—The original electromagnetic flow meters[1–5] utilized large magnets. An artery (A in Fig. 1a) was inserted into the gap between the pole pieces of the magnet. As the blood traversed the magnetic field at right angles, an emf was induced in the blood stream. This emf served as the measure of the rate of blood flow. The flow signal was picked up by means of two electrodes E_1 and E_2 touching the outer wall of the artery at the end points of a diameter perpendicular to the magnetic field, as shown in Figure 1*a*. The use of a constant magnetic field necessitated the use of nonpolarizable electrodes. The size of the magnet and the complication of using nonpolarizable electrodes restricted the application of this method to exteriorized arteries of anesthetized animals. The introduction of the alterating magnetic field[3, 6] simplified the design of the amplifying system and made the use of ordinary metal electrodes possible. This paved the way for the development of a small flow meter which could be implanted into animals to study blood flow in the conscious state in chronic experiments.[7] Chronic implantations up to 4 weeks' duration have been thus obtained. But the implanted units of the original design whose iron and copper skeleton weighed 5.5 gm and whose weight when encased in a plastic body was somewhat over 10 gm were still too large in comparison with the artery diameter for successful permanent implantations (Fig. 2*b*). The present paper describes simple designs of flow meter implants which, for a given artery diameter, are greatly reduced in size as compared to units built according to the original pattern.[7] Reductions in weight by a factor of twenty and, in some instances, more have been achieved. Such flow meters can be easily constructed to accommodate arteries of diameters in the neighborhood of 1 mm. The same objective of design can, of course, be used also for larger blood vessels. A special design for arteries over 1 cm in diameter is described below. Figure 2 shows a comparison between the original "miniature" flow meter implant[7] for an artery diameter of

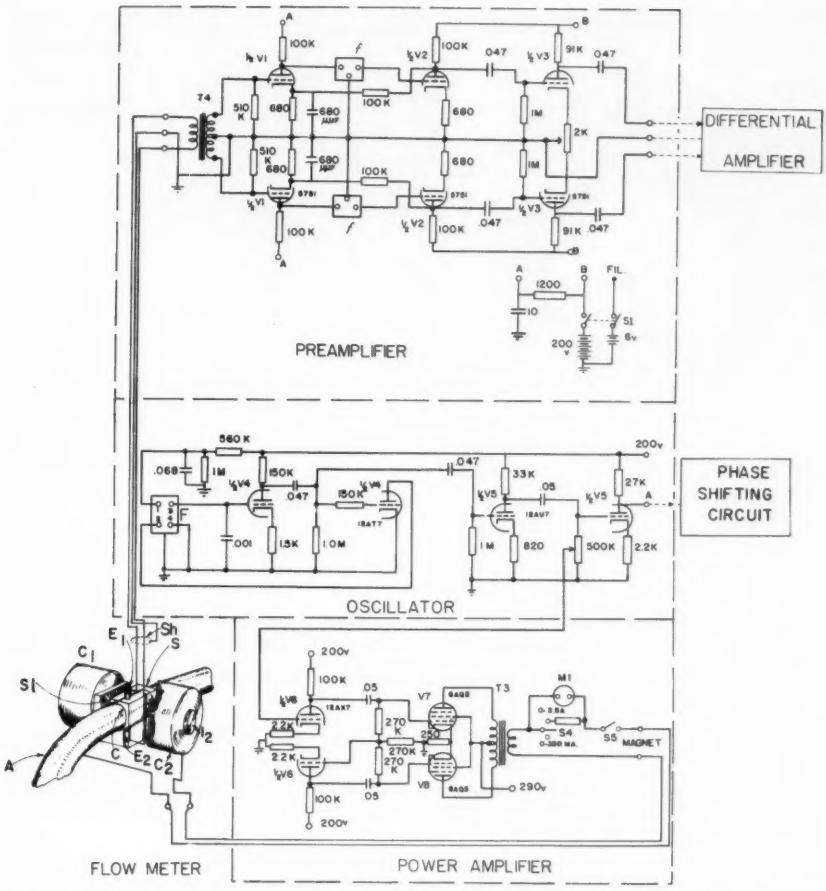

FIG. 1*a*.—Electromagnetic flow meter and associated electronic circuit, Part 1. The electronic components are represented by conventional symbols and their magnitudes are marked.

Flow meter section: I_2, magnet core; E_1, E_2, electrodes; C_1, C_2, magnet coils; S, sleeve; C, sleeve channel; Sl, slit; A, artery; Sh, grounded shield.

Power amplifier section: M_1, two-range ammeter; S_4, range switch; S_5, magnet switch; T_3, output transformer (Triad S55X).

Oscillator section: F, Tuning fork (Philamon Labs N. Y. type MJT400).

Preamplifier section: T_4, input transformer (triad geoformer g40) T.S; f, 400 cps LC filters (UTC type BMI 400).

2 mm (*b*) and the subminiature unit to be described below (*a*) weighing approximately 0.5 gm for an artery of 1.5 mm.

Illustrations of Blood Flow Changes Observed with Miniaturized Flow Meter Implants.—Figure 3*b–h* shows typical records of blood flow obtained with a transducer of the type shown in Figure 2*a* in a carotid artery of a conscious cat over a period of about 4 months. The base lines (indicated by arrows) were obtained by switching off the magnet. Figure 3*b* shows the contour of the blood flow during the cardiac cycle taken at high record speed. The following tracings, taken at lower

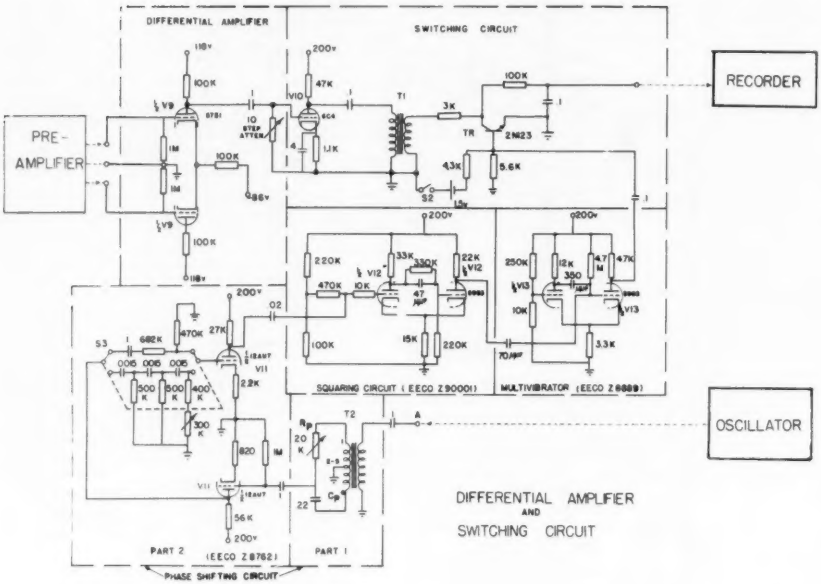

FIG. 1*b*.—Electronic circuit, Part 2.

Differential amplifier section: 10-step attenuator changes amplification by a factor of 2 per step.

Switching circuit: T_1, interstage transformer (Triad A42Z); S_2, switch; TR, switching transistor (Gen. Electr. 2N 123).

Recorder: SANBORN "Polyviso" model 67-1200 with model 67-300 dc amplifier.

Squaring ciacuit: plug-in unit EECO Z 90001.

Multivibrator: plug-inunit EECO Z 8889.

Phase shifting circuit: Part 1, T_2, transformer (Triad A4X); Rp, phase control rheostat with dial; Cp, capacitor.

Part 2, C_3, phase control switch (dpdt).

Variation of Rp permits a continuous change in the phase of the switching signal applied to TR over a range of approximately 150°; flipping of switch S_2 introduces an abrupt 90° shift in the phase of the switching signal applied to TR.

speeds, show changes in carotid blood flow in a conscious cat exposed to various sensory and emotional stimuli (see legends). The possibility of using this method for studies of effects of emotional stress on blood circulation through various organs is obvious. Of special interest is the possibility of taking simultaneous recordings from implants about arteries supplying different organs for studies of redistribution of blood flow in response to various physiological and psychological stimuli. Experiments of this kind are now in progress.

Miniaturization of the Flow Meter Design.—The possibility of designing symmetrical tuned a-c amplifiers (Figs. 1*a* and 1*b*) capable of yielding an adequately measureable output, with input signals of the order of 1μV at a noise level not exceeding 0.2 μV, makes it practical to sacrifice field strength and efficiency of the magnet in favor of a greatly simplified design offering maximum ease of construction and a minimum of irritation in the animal's body by reduction in size of the implant. With an artery of 1.5 mm diameter, a magnetic field of 200 gauss will yield a flow signal of approximately 17 μV for a flow of 1 cc/sec. The design described below corresponds approximately to these specifications.

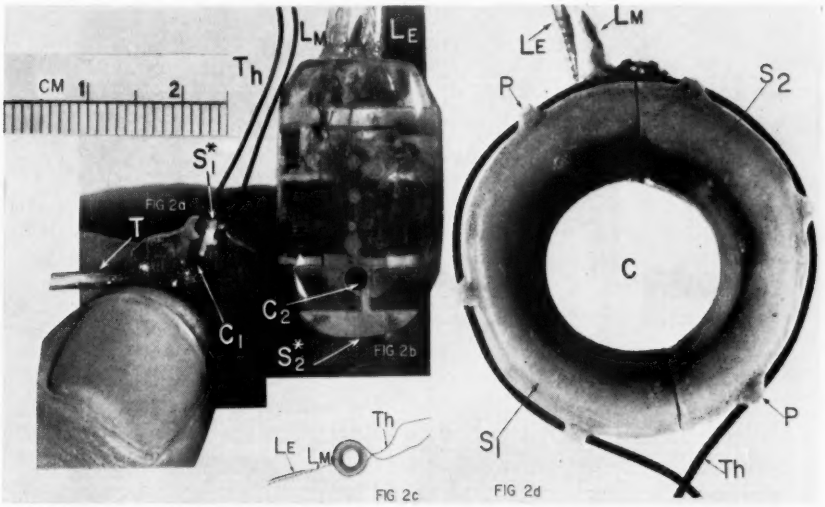

FIG. 2.—Comparison between different designs of implantable blood flow meters.

(2*a*) Subminiature unit for an artery 1.5 mm in diameter. T, polyvinyl chloride tubing containing the leads; C_1, channel for insertion of the blood vessel; S^*_1, shutter closing the slit through which the blood vessel is inserted into the channel C_1; Th, thread which facilitates handling the shutter S^*_1.

(2*b*) The original design (ref. 7) for a 2-mm blood vessel. L_M, magnet leads; L_E, electrode leads; C_2, channel for blood vessel; S^*_2, shutter.

(2*c*) The coreless flow meter, shown in Figure 2*d* in proper proportion in relation to the units depicted in Figures 2*a* and 2*b*, is shown reduced in size to match the diameter of channel C_1. L_E, electrode leads; L_M, magnet leads; Th, thread.

(2*d*) Flow meter for a blood vessel of 1.5-cm. diameter (shown reduced $^1/_{10}$ in Fig. 2*c*) utilizing no iron core. The bent coreless coils are sealed in plastic material in the two sections S_1 and S_2 which can be separated; the section S_1 contains both electrodes; C is the channel for the blood vessel formed by uniting the sections S_1 and S_2; L_E, electrode leads; L_M, magnet leads; Th, thread which is tied to prevent the sections S_1 and S_2 from coming apart after insertion of the blood vessel; P, plastic protrusions with perforations through which the thread Th is passed.

The flow transducer skeleton, shown in Figure 4*a*, consists of 3 main parts, C_1, C_2, and S. C_1 and C_2 are coils generating the magnetic field. They are connected in series through the terminals W_2 and W_3 which are grounded through wire G. Each coil consists of 200 turns of AWG #36 teflon insulated copper wire wound on an iron core I.† The ends of the cores I_1 and I_2 are connected by a grounded wire J (only the upper connection is shown). A current of 200 mA (rms) is used for continuous operation yielding a magnetic field of approximately 200 gauss peak value. 400 mA may be used for short periods of time in intermittent operation.

S is the lucite "sleeve" into whose channel C the blood vessel is slipped through the slit S1 (the width of S1 should be between $^1/_4$ and $^1/_3$ of the artery diameter). The electrodes E_1 and E_2 (0.5 mm gold wires) which are filed flush with the wall of channel C, establish the contact with the vessel. The flow signal picked up by the electrodes is conveyed by the leads L_1 and L_2 to the amplifying and recording system. The lead L_2 runs through a thin, shallow groove (not shown in the figure) to the electrode E_2. It is desirable, in order to minimize the "transformer emf" (see below), that the plane passing through this groove and the center axis of the elec-

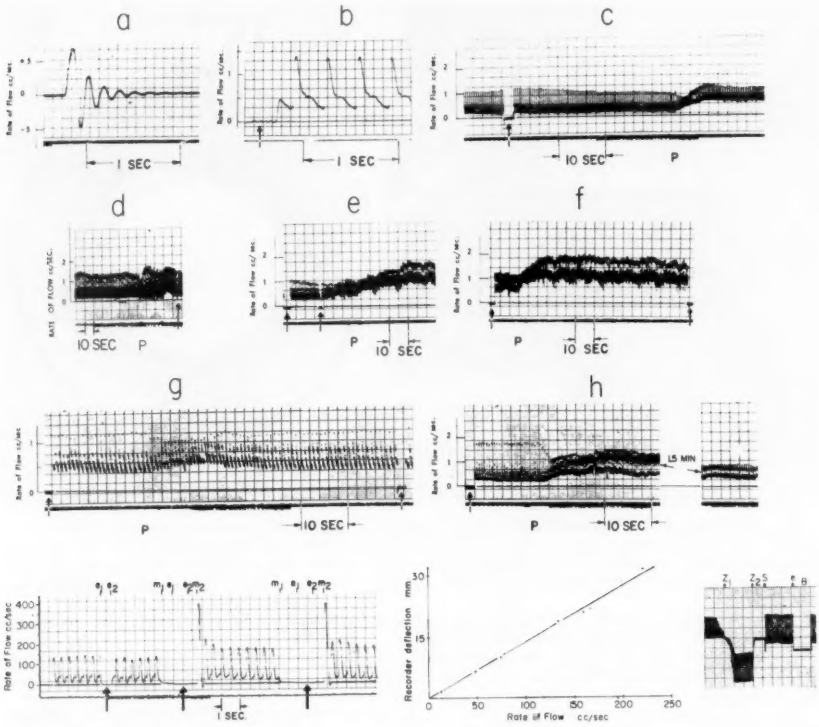

FIG. 3.—Examples of flow records.

(3*a*) Illustration of the ability of the flow meter to record rapidly varying flow in magnitude as well as in direction. This record depicts a damped oscillation of a liquid column in a conduit.

The following records *c*–*h* show changes in the rate of blood flow through the right carotid artery of a conscious cat in response to various sensory stimuli. The freedom of motion of the cat is limited only by the lead wires 10 ft in length. The arrows indicate base-lines ascertained by switching off the magnet; the continuous base-lines have been drawn in accordance with these sample base-lines. The records shown have been taken over a period of approx. 4 months following the implantation.

(3*b*) Normal variation of blood flow throughout the cardiac cycle.

(3*c*) Exposure of cat to ammonia fumes at point P.

(3*d*) Sleeping cat is awakened at P by a tactile stimulus.

(3*e*) Increase in systolic as well as diastolic blood flow in response to the odor of catnip presented at P.

(3*f*) Response to presentation of food at P and subsequently increased blood flow during the process of eating. It is noteworthy that immediately following P, the relative increase in diastolic flow is much greater than the relative increase in systolic flow.

(3*g*) Transient drop in systolic flow coupled with an increase in diastolic flow immediately following a startling light stimulus (a photographic flash exposure close to the cat's eyes at P).

(3*h*) Fright reaction induced by exposure of the cat to a dog at P. There is a sharp drop in systolic blood flow coupled with a rise in diastolic blood flow and an increase in heart rate. The last brief record section, taken two minutes after the fright stimulus, shows a persistent notable reduction in systolic blood flow.

(3*i*) Blood flow in the descending thoracic aorta of an anesthetized dog (27 kg) obtained with a flow meter of the design shown in Figures 2*d* and 4*c*. The arrows indicate base-lines obtained by switching the magnet current off (at e_1) and on (at e_2). The record also shows the close agreement between the base-lines thus obtained and the determination of the base-line by clamping the artery off (at m_1) and releasing it (at m_2). The compression interval between m_1 and m_2 is approximately 2 sec. The first systolic blood flow pulse (reinforced by retouching) following the occlusion of the artery is about three times as large as the preceding normal ones and is followed by blood flow notably above the normal level.

(3*j*) Calibration graph showing instrument reading as a function of flow rate for a flow meter of the type shown in Figures 2*d* and 4*c*.

(3*k*) Illustration of the method of adjusting the switching phase (see text).

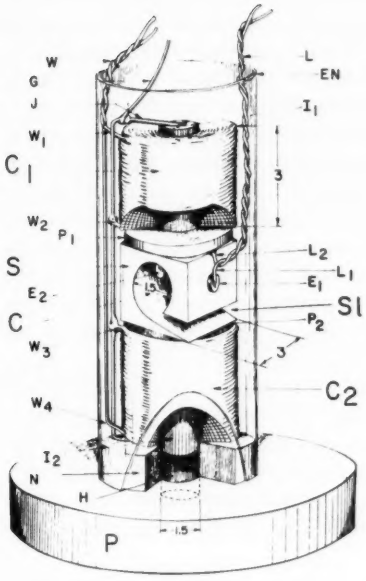

FIG. 4a

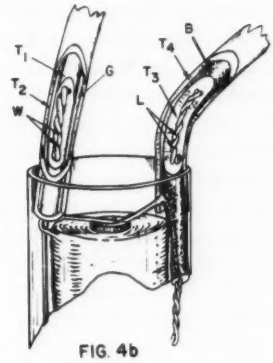

FIG. 4b

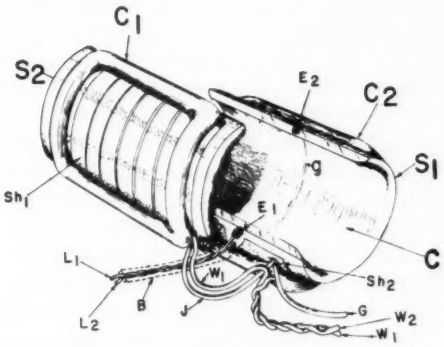

FIG. 4c

FIG. 4.—Flow meter designs in which the volume occupied by the magnet is greatly reduced as compared to the volume occupied by the artery.

(4*a*) Subminiature flow meter for a 1.5 mm artery. (The dimensions are given in the diagram in mm.) The unit is shown with accessories used for its fabrication. C_1 and C_2, magnet coils; S, sleeve; P, pedestal; N, pedestal neck; H, hole; I_1, I_2, iron cores of coils C_1, C_2; E_1, E_2, electrodes; Sl, slot; P_1, P_2, lucite mounting plates; L_1, L_2, electrode leads; W_1, W_2, W_3, W_4, coil terminal wires; J, wire joining the cores I_1 and I_2; G, ground lead; W, magnet leads; L, electrode leads; EN, cellulose envelope serving as a mold in the process of casting.

(4*b*) Shielding and insulation of the leads. The symbols for the wires have the same meaning as in Figure 4*a*. T_1, T_2, T_3, and T_4, polyvinyl chloride tubing. B, grounded wire braid shielding the electrode leads.

(4*c*) Coreless flow meter, especially suitable for large arteries. S_1, S_2, two sections which, when united, present the aspect shown in Figure 2*d* (after coating the coils with plastic material and insulating the wires with polyvinyl chloride tubing). The section S_1 contains the electrodes E_1 and E_2. C_2, magnet coils; C, channel for artery; g, groove for electrode lead L_2 (on the convex side of section S_1); Sh_1, Sh_2, silver shields [the "ribs" of Sh_1 (see text) can be seen framed by coil C_1]; L_1, L_2, electrode leads; W_1, W_2, magnet coil leads; G, ground lead; J, grounded wire joining the shields Sh_1 and Sh_2; B, grounded wire braid shielding the electrode leads L_1 and L_2. The insulation of the leads and their attachment to the flow meter body are as described in the text in connection with Figure 4*b*.

trodes be as nearly as possible parallel to the magnetic field lines. It is also important for the electrode axis to be as nearly as possible perpendicular to the axis of channel C to minimize pick-up due to possible axial currents passing through channel C.

P_1 and P_2 are lucite centering plates cemented to the sleeve S. Protrusions of the cores I_1 and I_2 fit snugly into central holes provided in these plates. The axes of

the cores are thus aligned so as to pass through the electrode axis intersecting it at right angles. The coils and sleeve are assembled as shown in Figure 4*a* with the bottom protrusion of Core I_2 inserted into the hole H in the neck N of the pedestal P. The channel C and slit S1 are filled with Wood's metal prior to mounting the assembly on the pedestal. A cylindrical envelope EN of cellulose adhesive tape is wrapped around the neck N of the pedestal surrounding the assembly, as shown in Figure 4*a*. The nonadhesive surface of the tape is turned inward. The assembly is now ready to be cast in a nonirritant plastic material. "Teets denture material" (TDM) has been used successfully for this purpose. The polymerizing liquid is poured into the mold shown in Figure 4*a* which is placed into a vacuum desiccator. Evacuation for about 2 minutes removes air from small crevices so that on re-establishment of atmospheric pressure, the monomer is forced into the interspaces between the coil wires and other crevices. After completion of the polymerization, within about a half hour, the casting may be hardened by curing for about a half hour at 100°C.

The final step consists of insulating the lead wires by insertion into polyvinyl chloride tubing. It is most important to secure a watertight seal of this tubing to the plastic body of the flow transducer. The ends of the tubing are submerged into the polymerizing liquid which is poured into the mold over the upper coil C_1 so as to cover the protrusion of the core I_1. The tube ends are pre-soaked in the monomer for about one hour before submersion in TDM. This secures a leak-proof bond.

The arrangement of wires and insulators is shown in Figure 4*b*. The symbols have the same meaning as in Figure 4*a*. T_1, T_2, T_3, and T_4 are thin polyvinyl chloride tubes.‡ B is a thin wire braid which shields the electrode leads L. It is grounded by connection (not shown in the figure) to the ground lead G. The terminals of the wires W, L, and G are attached to small plugs. A gold plated-wire connected to ground wire G is permitted to emerge from the interior of the transducer to the surface for the purpose of grounding the animal.

The shape of the slit is seen in its initial stage in Figure 4*a*. The final shape is determined by the teflon shutter S_1* shown in Figure 2*a*. After the flow meter is removed from the mold, some plastic material is filed away to expose the Wood's metal. The latter is removed from C and S1 by melting. The lower portion of the prefabricated shutter S_1* is then inserted into S1 and the shutter is completely covered with TDM paste. After hardening of the paste, a sliding bed, fitting exactly the contour of the shutter, is obtained in which the shutter slides freely in and out. The thread Th facilitates its handling.

The flow meter is completed by depositing TDM paste at the lower end of the assembly to insulate the tip of the iron core I_2. After the plastic material has become hard, the unit is filed to the desired shape and polished and the sharp edges at the ends of the channel C are rounded. It is recommended to coat the outer flow meter surface and the adjacent polyvinyl tubing with a coat of a 10 per cent solution of polyvinyl chloride in cyclo-hexanone.

Coreless Flow Meter for Large Arteries.—A similar principle of design can be used for large transducers. The large flow in major blood vessels permits the reduction of the magnetic field to the order of magnitude of 10 gauss S.[§] In this case "miniaturization" consists in greatly reducing the volume of the magnet in com-

parison to the volume of the artery channel C. Figure 4*c* shows the skeleton of a flow meter, using no iron core, for an artery of 1.5 cm diameter. Figure 2*d* shows the end-on view of a flow meter of this type in correct proportion as compared to Figures 2*a* and *b*. The bent coils C_1 and C_2, using only 50 turns per coil, are cemented directly to the sleeve. Gauge #29 teflon insulated wire is used at 1 amp in continuous operation. A field of 20 gauss is obtained at the center of C. Figure 3*j* shows the linear calibration obtained with a unit of this type and Figure 3*i* shows normal flow tracings taken in the descending thoracic aorta of an anesthetized 27 kg dog. The baselines marked by arrows were secured by switching off the magnet. The two last baselines were taken so as to compare the effects of occluding the artery with switching off the magnet (see legend). Both base lines are seen to coincide as expected. The large increase in blood flow following an occlusion of the aorta for a little over 2 seconds is noteworthy.

To depict the reduction in space occupied by the magnet in this unit, as compared with the designs shown in Figures 2*a* and *b*, the aperture of the unit shown in Figure 2*d* has been reduced to 1.5 mm in diameter as shown in Figure 2*c*.

The Electronic Circuit.—The important step of ascertaining the base line corresponding to zero flow without interruption of flow is accomplished by special circuit design rather than through special adjustment of the flow meter sleeve previously described.[8] Phase sensitive rectification of the sinusoidal amplifier output is accomplished by a transistor switching circuit (Fig. 1*b*) which is triggered by rectangular voltage pulses controlled by a phase-shifting network. The latter is adjusted so that the transistor switch TR is open at the moment when the flow signal which is in phase with the magnetic field, is maximal. At this moment the "transformer emf" induced by the changing magnetic field in the loop formed by lead L_2 which runs from electrode E_2 toward electrode E_1 (Fig. 4*a*) is zero. Thus, this adjustment secures optimum rectification of the flow signal and rejection of the transformer emf which is in phase quadrature with respect to the flow signal. This method of rectification with phase discrimination yields an output which reverses direction with reversal of flow. This property is illustrated by the record of damped oscillations of alternating fluid flow in a conduit in Figure 3*a*. The rejection of the transformer emf is illustrated in Figure 3*i* where the base line does not change when the magnet is switched off after stopping the blood flow by clamping the artery.

Rectification of flow signals by means of a somewhat more complex gating circuit using exclusively vacuum tubes has been described previously in connection with flow meters actuated by square wave current pulses.[9] The present circuit, utilizing a transistor switch, accomplishes the same purpose more simply with sinusoidal currents energizing the magnet. The first suggestion of use of a gating circuit with an electromagnetic flow meter appears to have been made by James.[10]

The various sections of the electronic circuit shown in Figures 1a and 1*b* interact as follows. The power amplifier shown in Figure 1*a* supplies the 400 cps current to the magnet. It derives its input signal from the oscillator section which also supplies the synchronizing signal to the switching circuit. The flow signal is derived from the electrodes E_1 and E_2 of the flow meter and is conveyed to the input of the preamplifier whose input is connected to the differential amplifier of Figure 1*b*. The output signal of the latter passes through the switching circuit controlled by

the transistor switch TR. The rectified output of this circuit reaches the pen recorder (SANBORN "Polyviso" Model 67-1200).

The remainder of the circuit shown in Figure 1*b* serves the important purpose of delivering properly synchronized voltage pulses to the transistor switch TR to insure that the flow signal is sampled at the moment when the transformer emf is zero. The switching pulses, lasting about $^1/_{20}$ of a cycle, are derived from the multivibrator which is triggered by pulses derived from the squaring circuit. The phase of the output of the squaring circuit is controlled by the signal fed into it from the phase-shifting circuit which, in turn, receives its input signal from the oscillator section (Fig. 1*a*). The phase-shifting circuit consists of two sections. The first one permits a continuous variation in phase over a range of about 150° by control of the rheostat Rp which is in series with the capacitor Cp. The second part permits the introduction of an abrupt 90° jump in phase in the signal delivered to the squaring circuit through flipping the d.p.d.t. switch S_3. The continuous and discon tinuous phase shifts are used as follows in adjustment of the instrument for rejection of the transformer emf. We make use of the fact that the arterial blood flow is pulsating. Thus, when we adjust the resistor Rp so as to reduce the pulsations in our signal to zero, we know that we are sampling our flow meter output at the most inappropriate moment, namely, when the pulsating flow signal is zero and the transformer emf is at its maximum. Flipping the switch S_3 shifts the switching phase through 90° and thus reverses the situation. The switch TR transmits now the flow signal to the recorder at the maximum point of the signal at which time the transformer emf is zero. This procedure of adjustment is illustrated in Figure 4*k*. As the resistor Rp of the phase shifting circuit (Part I, Fig. 1*b*) is varied, a point Z_1 is reached at which the recorded upward pulses, caused by pulsating blood flow, vanish. When this point is passed, the recorded pulses are directed downward due to reversal of the switching phase of the transistor TR(Fig. 1*b*). The point Z_1 determines the dial setting of Rp for elimination of the flow signal. When Rp is left at the setting corresponding to Z_1, a line Z_2 is recorded whose height above the baseline B measures the transformer emf. At S the switch S_3 of Fig. 1*b* is flipped and the transistor switch TR is phased to sample the flow signal and reject the transformer emf. The blood flow record between points S and e is thus obtained. At e the magnet is turned off and the baseline B is obtained. At the end of the section B, the magnet is turned on again.

The flow meter design and circuitry described herein extend the applicability of the electromagnetic flow meter to blood vessels down to the size of 1 mm in diameter and possibly to smaller ones, making it possible to tackle problems on localized regional blood flow in conscious animals which could not be attempted heretofore because of the large bulk of the implant.

Summary.—A high gain low noise amplifying system coupled with a properly synchronized switching circuit makes it possible to minimize the bulk of the magnet of an electromagnetic flow meter. Designs are described permitting blood flow measurements in arteries ranging from about 1 mm to over 15 mm in diameter. The magnet is energized with 400 cps ac. The base-line is established by switching off the magnet. Illustrations of blood flow changes in a conscious, freely moving animal in response to sensory stimuli are given.

* This work has been supported by a grant from the Office of Naval Research.

† Nonlaminated round iron rods can be used for magnets of the small size shown here. For artery diameters above 2 mm, laminated cores of grain-oriented silicon steel are recommended.

‡ Tubing of greatly reduced diameter can be easily obtained from available tubing by stretching and heating it in an oven to approximately 120°C. for about a half hour.

§ For instance, for an aorta of 2 cm diameter and an average linear fluid velocity of 50 cm/sec, a field of 10 gauss will yield a flow signal of 10 μV.

[1] Kolin, A., *Proc. Soc. Exp. Biol. & Med.* **35,** 53 (1936).

[2] *Ibid.*, A., *Am. J. Physiol.* **119,** 355 (1937).

[3] *Ibid.*, **122,** 797 (1938).

[4] Wetterer, E., *Z. Biol.* **98,** 26 (1937).

[5] *Ibid.*, **99,** 158 (1938).

[6] Kolin, A., *Proc. Soc. Exp. Biol. & Med.*, **46,** 235 (1941).

[7] Kolin, A., N. S. Assali, G. Herrold, and R. Jensen, these PROCEEDINGS, **43,** 527 (1957).

[8] Kolin, A., G. Herrold, and N. S. Assali, *Proc. Soc. Exp. Biol. & Med.* **98,** 550 (1958).

[9] Denison, A. B., and M. P. Spencer, *Rev. Sc. Instr.* **27.** 707 (1956).

[10] James, W. G., *Rev. Sc. Instr.*, **22,** 989 (1951).

INFORMATION TO CONTRIBUTORS

THE PROCEEDINGS is the official organ of the NATIONAL ACADEMY OF SCIENCES and of the NATIONAL RESEARCH COUNCIL for the publication of brief accounts of important current researches of members of the ACADEMY and of the COUNCIL and of other American investigators. The PROCEEDINGS will aim especially to secure prompt publication of original announcements of discoveries and wide circulation of the results of American research among investigators in other countries and in all branches of science.

ARTICLES should be brief. The viewpoint should be comprehensive in giving the relation of the paper to previous publications of the author or of others and in exhibiting, where practicable, the significance of the work for other branches of science. Elaborate technical details of the work and long tables of data should be avoided, but authors should be precise in making clear the new results and should give some record of the methods and data upon which they are based.

MANUSCRIPTS should be prepared with a current number of THE PROCEEDINGS as a model in matters of form, and should be typewritten in duplicate with double spacing, the author retaining one copy. Illustrations should be confined to text figures of simple character, though more elaborate illustration may be allowed in special instances to authors willing to pay for their preparation and insertion. Particular attention should be given to arranging tabular matter in a simple and concise manner.

MANUSCRIPTS will be accepted only from members of the Academy who will assume responsibility for the propriety and scientific standards of the paper and for any printing costs in excess of those allowed them by the Academy. Manuscripts should be addressed to The PROCEEDINGS OF THE NATIONAL ACADEMY OF SCIENCES, 2101 Constitution Ave., Washington 25, D. C.

GALLEY PROOF will be sent, with the understanding that charges for author's corrections shall be billed to him. Authors should therefore make final revisions on the typewritten manuscripts. Original artwork or figures will be returned if requested when the manuscript is submitted. Because of time limitations, page proofs will not be sent.

REPRINTS should be ordered when proof is returned, on the reprint order form which will be enclosed with the proof. They will be furnished at cost, approximately as follows:

SCHEDULE OF REPRINT RATES (POSTAGE NOT INCLUDED)

Quantity	*up to 4 pp.*	*up to 8 pp.*	*up to 12 pp.*	*Standard Covers*
50	$ 7.85	$15.10	$22.30	$6.90
100	8.95	17.30	25.60	8.40
150	10.10	19.50	28.95	9.95
add'l 100's	2.20	4.40	6.60	3.05

CONTENTS

Page

GEOPHYSICS.—SYMPOSIUM ON SCIENTIFIC EFFECTS OF ARTIFICIALLY INTRODUCED RADIATIONS AT HIGH ALTITUDES 1141

CHAIRMAN'S INTRODUCTORY REMARKS . . . . . *By Richard W. Porter* 1141

THE ARGUS EXPERIMENT . . . . . *By N. C. Christofilos* 1144

SATELLITE OBSERVATIONS OF ELECTRONS ARTIFICIALLY INJECTED INTO THE GEOMAGNETIC FIELD . . . . . *By James A. Van Allen, Carl E. McIlwain, and George H. Ludwig* 1152

PROJECT JASON MEASUREMENT OF TRAPPED ELECTRONS FROM A NUCLEAR DEVICE BY SOUNDING ROCKETS . . . . . *By Lew Allen, Jr., James L. Beavers II, William A. Whitaker, Jasper A. Welch, Jr., and Roddy B. Walton* 1171

THEORY OF GEOMAGNETICALLY TRAPPED ELECTRONS FROM AN ARTIFICIAL SOURCE . . . . . *By Jasper A. Welch, Jr., and William A. Whitaker* 1190

OPTICAL, ELECTROMAGNETIC, AND SATELLITE OBSERVATIONS OF HIGH-ALTITUDE NUCLEAR DETONATIONS—PART I . . . . . *By Philip Newman* 1208

OPTICAL, ELECTROMAGNETIC, AND SATELLITE OBSERVATIONS OF HIGH-ALTITUDE NUCLEAR DETONATIONS—PART II . . . . . *By Allen M. Peterson* 1221

BIOCHEMISTRY.—ON CHARGE TRANSFER COMPLEXES BETWEEN SUBSTANCES OF BIOCHEMICAL INTEREST . . . . . *By Irvin Isenberg and Albert Szent-Györgyi* 1229

BIOCHEMISTRY.—ON THE ELECTRON PARAMAGNETIC RESONANCE OF ADENOSINE TRIPHOSPHATE . . . . . *By Irvin Isenberg and Albert Szent-Györgyi* 1232

BIOCHEMISTRY.—ENZYMES IN A GLYCOGEN STORAGE MYOPATHY . . . . . *By Joseph Larner and Carlos Villar-Palasi* 1234

BIOCHEMISTRY.—GLYCOGEN SYNTHESIS IN MUSCLE LACKING PHOSPHORYLASE . . . . . *By Rudi Schmid, Phillips W. Robbins, and Robert R. Traut* 1236

BIOCHEMISTRY.—HYDROXYMETHYLDEOXYCYTIDYLATE KINASE FORMATION AFTER BACTERIOPHAGE INFECTION OF *Escherichia coli* . . . . . *By Ronald Somerville, Kaney Ebisuzaki, and G. Robert Greenberg* 1240

BIOCHEMISTRY.—THE ROLE OF ORGANIC SUBSTRATES IN BACTERIAL PHOTOSYNTHESIS . . . . . *By R. Y. Stanier, M. Doudoroff, R. Kunisawa, and R. Contopoulou* 1246

BIOCHEMISTRY.—SUBCELLULAR PARTICULATE SYSTEMS AND THE PHOTOCHEMICAL APPARATUS OF *Rhodospirillum rubrum* . . . . . *By Alice L. Tuttle and Howard Gest* 1261

CHEMISTRY.—A CRITERION FOR ORBITAL HYBRIDIZATION AND CHARGE DISTRIBUTION IN CHEMICAL BONDS . . . . . *By Martin Karplus and D. M. Grant* 1269

CHEMISTRY.—THE EFFECTS OF SPREADING TECHNIQUE AND PURITY OF SAMPLE ON THE EVAPORATION RESISTANCE OF MONOLAYERS . . . . . *By Victor K. La Mer and Geoffrey T. Barnes* 1274

GENETICS.—THE FORMATION OF A NEW ENZYMATICALLY ACTIVE PROTEIN AS A RESULT OF SUPPRESSION . . . . . *By Irving P. Crawford and Charles Yanofsky* 1280

MATHEMATICS.—A MATHEMATICAL THEORY OF ADAPTIVE CONTROL PROCESSES . . . . . *By Richard Bellman and Robert Kalaba* 1288

MATHEMATICS.—PARTITIONS IN HOMOGENEOUS, FINITE ABELIAN GROUPS *By Eckford Cohen* 1290

MATHEMATICS.—ON ENTIRE SOLUTIONS OF NONLINEAR DIFFERENTIAL EQUATIONS . . . . . *By Raymond Redheffer* 1292

PATHOLOGY.—CRITICAL SIGNIFICANCE OF BREAST STRUCTURE IN THE INDUCTION OF MAMMARY CANCER IN THE RAT . . . . . *By Charles Huggins, Lorraine C. Grand, and Filomena P. Brillantes* 1294

PHYSICS.—MICROSCOPIC TESTS OF SYMMETRY PRINCIPLES . . . . . *By G. Feinberg and M. Goldhaber* 1301

PHYSIOLOGY.—MINIATURIZATION OF THE ELECTROMAGNETIC BLOOD FLOW METER AND ITS USE FOR THE RECORDING OF CIRCULATORY RESPONSES OF CONSCIOUS ANIMALS TO SENSORY STIMULI . . . . . *By Alexander Kolin and Raymond T. Kado* 1312